

On Noncoherent Fading Relay Channels at High SNR

Tobias Koch

University of Cambridge

Cambridge, UK

Email: `tobi.koch@eng.cam.ac.uk`

Gerhard Kramer

Technische Universität München

München, Germany

Email: `gerhard.kramer@tum.de`

Abstract

The capacity of noncoherent fading relay channels is studied, where the communication between the transmitter and the receiver is supported by a relay, where the links between the terminals are fading channels, and where all terminals are aware of the fading statistics, but not of their realizations. It is shown that if the link between the transmitter and the receiver supports higher communication rates than the link between the relay and the receiver, then at high SNR it is best to turn the relay off. It is further shown that if the link between the transmitter and the relay supports higher communication rates than the link between the relay and the receiver, then at high SNR one can achieve communication rates that are within one bit of the capacity of the multiple-input single-output fading channel that results when the transmitter and the relay can cooperate.

1 Introduction

We study the capacity of *fading relay channels*. A relay channel consists of a transmitter, a receiver, and a relay, which supports the transmitter in communicating with the receiver. The word “fading” refers to the variation in the strength of the links between these terminals. Coherent fading relay channels were studied, e.g., in [1], [2]. For such channels, the fading coefficients are available at the corresponding receiving terminals.

The assumption that the fading coefficients are available at the receiving terminals is commonly justified by saying that these coefficients vary slowly over time and can therefore be estimated by transmitting training sequences. However, this assumption yields overly-optimistic results, since it is *prima facie* not clear whether the fading coefficients can be estimated perfectly, and since the transmission of training sequences reduces the achievable communication rates (training sequences do not contain information). For instance, in the point-to-point case, where a transmitter communicates with a receiver (without the aid of a relay), the loss in not knowing the fading coefficient at the receiver beforehand can be substantial. Indeed, if the fading is *regular* in the sense that the present fading coefficient cannot be predicted perfectly from its infinite past, then, at high signal-to-noise ratio (SNR), the capacity grows double-logarithmically with the SNR [3], which is in stark contrast to the logarithmic growth in the coherent case [4]. If the fading is *nonregular* in the sense that the present fading coefficient can be predicted perfectly from its past, then the capacity can grow logarithmically with the SNR, but the pre-log, defined as the limiting ratio of capacity to $\log \text{SNR}$ as SNR tends to infinity, depends on the fading’s autocovariance function and is typically strictly smaller than one [5].

In this paper, we study the capacity of noncoherent fading relay channels with regular fading. For such channels the terminals are aware only of the laws of the fading coefficients, but not of their realizations. We derive two basic results. First, we show that if the link between the transmitter

The material in this paper was presented in part at the 2005 IEEE International Symposium on Information Theory (ISIT), Adelaide, Australia. T. Koch received funding from the European Community’s Seventh Framework Programme (FP7/2007-2013) under grant agreement No. 252663. G. Kramer was supported in part by the Board of Trustees of the University of Illinois Subaward No. 04-217 under National Science Foundation Grant CCR-0325673, and by an Alexander von Humboldt Professorship endowed by the German Federal Ministry of Education and Research.

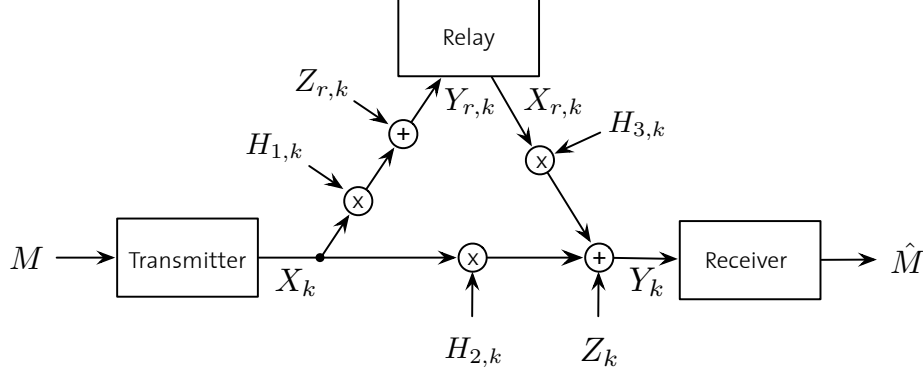


Figure 1: The relay channel.

and the receiver supports higher communication rates than the link between the relay and the receiver, then at high SNR it is optimal to turn the relay off. Second, we show that if the link between the transmitter and the relay supports higher communication rates than the link between the relay and the receiver, then at high SNR one can achieve communication rates that are within one bit of the capacity of the multiple-input single-output (MISO) fading channel that results when the transmitter and the relay can cooperate. Thus, at high SNR the rate penalty for establishing cooperation between the transmitter and the relay is not greater than one bit.

Note that we model the fading coefficients as stationary and ergodic stochastic processes whose autocovariance functions determine the fading's time-variation. This excludes the so called *block-fading model* introduced by Marzetta and Hochwald [6]. (The block-fading model is not stationary.) It turns out that, in the point-to-point case, the block-fading model and the stationary and ergodic fading model yield completely different capacity behaviors at high SNR, cf. [7] and [3, 5]. We expect that this is also the case for fading relay channels.

This paper is organized as follows. Section 2 describes the mathematical channel model. Section 3 introduces channel capacity and defines the fading number. Section 4 presents the main results. Section 5 presents nonasymptotic bounds on the capacity of the relay channel. Section 6 contains the proof of the upper bound (Theorem 1), and Section 7 contains the proof of the lower bound (Theorem 3). Sections 8 and 9 conclude the paper with a discussion and summary of the obtained results.

2 Channel Model

The relay channel, depicted in Figure 1, consists of three terminals: the transmitter, the receiver, and the relay. The message M to be transmitted over the relay channel is assumed to be uniformly distributed over the set $\mathcal{M} = \{1, \dots, |\mathcal{M}|\}$, where $|\mathcal{M}|$ is a positive integer. The transmitter maps M to the length- n sequence $X_1^n = X_1, \dots, X_n$, where n is referred to as the *blocklength*. Thus, $X_1^n = \phi_n(M)$ for some mapping $\phi_n: \mathcal{M} \rightarrow \mathbb{C}^n$ (where \mathbb{C} denotes the set of complex numbers). At each time instant $k \in \mathbb{Z}$ (where \mathbb{Z} denotes the set of integers), the relay observes $Y_{r,k} \in \mathbb{C}$ and emits the symbol $X_{r,k} \in \mathbb{C}$, which is a function of the previously received symbols $Y_{r,1}^{k-1}$, i.e., $X_{r,k} = \varphi_{n,k}(Y_{r,1}^{k-1})$, $k = 1, \dots, n$ for some mapping $\varphi_{n,k}: \mathbb{C}^{k-1} \rightarrow \mathbb{C}$. The receiver observes the channel output symbols Y_1^n from which it guesses M . The receiver's guess is denoted by \hat{M} , i.e., $\hat{M} = \psi_n(Y_1^n)$ for some mapping $\psi: \mathbb{C}^n \rightarrow \mathcal{M}$.

The time- k channel outputs $Y_{r,k}$ and Y_k corresponding to the channel inputs x_k and $x_{r,k}$ are given by

$$Y_{r,k} = H_{1,k}x_k + Z_{r,k}, \quad k \in \mathbb{Z} \quad (1)$$

$$Y_k = H_{2,k}x_k + H_{3,k}x_{r,k} + Z_k, \quad k \in \mathbb{Z}. \quad (2)$$

Here $\{H_{1,k}, k \in \mathbb{Z}\}$, $\{H_{2,k}, k \in \mathbb{Z}\}$, $\{H_{3,k}, k \in \mathbb{Z}\}$, $\{Z_{r,k}, k \in \mathbb{Z}\}$, and $\{Z_k, k \in \mathbb{Z}\}$ are stationary and ergodic stochastic processes that take on values in \mathbb{C} and are independent of each other. Fur-

thermore, $\{H_{1,k}, k \in \mathbb{Z}\}$ and $\{Z_{r,k}, k \in \mathbb{Z}\}$ are of a joint law that does not depend on $\{x_k, k \in \mathbb{Z}\}$, and $\{H_{2,k}, k \in \mathbb{Z}\}$, $\{H_{3,k}, k \in \mathbb{Z}\}$, and $\{Z_k, k \in \mathbb{Z}\}$ are of a joint law that does not depend on $\{(x_k, x_{r,k}), k \in \mathbb{Z}\}$.

The additive noise terms $\{Z_k, k \in \mathbb{Z}\}$ and $\{Z_{r,k}, k \in \mathbb{Z}\}$ are both sequences of independent and identically distributed (i.i.d.), zero-mean, circularly-symmetric, complex Gaussian random variables of variance σ^2 . The multiplicative noise terms (“fading”) $\{H_{1,k}, k \in \mathbb{Z}\}$, $\{H_{2,k}, k \in \mathbb{Z}\}$, and $\{H_{3,k}, k \in \mathbb{Z}\}$ are zero-mean, unit-variance, stationary and ergodic, circularly-symmetric, complex Gaussian processes with the respective spectral distribution functions $F_1(\cdot)$, $F_2(\cdot)$, and $F_3(\cdot)$. Thus, $F_\ell(\cdot)$, $\ell = 1, 2, 3$ are bounded and nondecreasing functions on $[-1/2, 1/2]$ satisfying

$$\mathbb{E}[H_{\ell,k+m}H_{\ell,k}^*] = \int_{-1/2}^{1/2} e^{i2\pi m\lambda} dF_\ell(\lambda), \quad \ell = 1, 2, 3 \quad (3)$$

where $i = \sqrt{-1}$. We assume a *noncoherent* channel model where the transmitter, receiver, and relay are not aware of the realization of the fading processes $\{H_{\ell,k}, k \in \mathbb{Z}\}$, $\ell = 1, 2, 3$ but only of their joint law. We further assume that the fading processes $\{H_{\ell,k}, k \in \mathbb{Z}\}$, $\ell = 1, 2, 3$ are regular in the sense that they satisfy

$$\int_{-1/2}^{1/2} \log F'_\ell(\lambda) d\lambda > -\infty, \quad \ell = 1, 2, 3 \quad (4)$$

where $F'_\ell(\cdot)$ denotes the derivative of $F_\ell(\cdot)$. (Note that, since $F_\ell(\cdot)$ is monotonic, it is almost everywhere differentiable. At the discontinuity points of $F_\ell(\cdot)$ the derivative $F'_\ell(\cdot)$ is undefined.) This implies that the mean-square error in predicting $H_{\ell,0}$ from $H_{\ell,-1}, H_{\ell,-2}, \dots$, which is given by [8]

$$\epsilon_\ell^2 = \exp\left(\int_{-1/2}^{1/2} \log F'_\ell(\lambda) d\lambda\right) \quad (5)$$

is strictly positive. We also have $\epsilon_\ell^2 \leq 1$, $\ell = 1, 2, 3$ since we take $\{H_{\ell,k}, k \in \mathbb{Z}\}$ to have unit variance. Roughly speaking, we can thus say that a regular process cannot be predicted perfectly from its past.

We assume that the channel inputs X_k and $X_{r,k}$ satisfy a *peak-power constraint*, i.e., we have with probability one

$$|X_k|^2 \leq A^2, \quad k \in \mathbb{Z} \quad (6)$$

and

$$|X_{r,k}|^2 \leq A_r^2, \quad k \in \mathbb{Z} \quad (7)$$

for some positive real A and A_r . We assume that

$$A_r = \rho A \quad (8)$$

for some $\rho > 0$ (independent of A), and we define the SNR as

$$\text{SNR} \triangleq \frac{A^2}{\sigma^2}. \quad (9)$$

Note that the results presented in this paper continue to hold if the peak-power constraints are replaced by average-power constraints.

3 Channel Capacity and Fading Number

A rate $R(\text{SNR})$ (in nats per channel use) is said to be *achievable*, if for every $\delta > 0$ there exist mappings ϕ_n , $(\varphi_{n,1}, \dots, \varphi_{n,n})$, and ψ_n satisfying (6) and (7) such that

$$\frac{\log |\mathcal{M}|}{n} > R(\text{SNR}) - \delta$$

and such that the error probability $\Pr(\hat{M} \neq M)$ tends to zero as n tends to infinity. (Here $\log(\cdot)$ denotes the natural logarithm function.) The *capacity* $C(\text{SNR})$ is defined as the supremum of all achievable rates.¹

We will focus on the asymptotic behavior of capacity at high SNR. In the point-to-point case, Lapidoth and Moser demonstrated that for regular fading, the capacity satisfies [3, Th. 4.2]

$$\overline{\lim}_{\text{SNR} \rightarrow \infty} \{C(\text{SNR}) - \log \log \text{SNR}\} < \infty \quad (10)$$

where $\overline{\lim}$ denotes the *limit superior*. They defined the *fading number* χ as [3, Def. 4.6]

$$\chi \triangleq \overline{\lim}_{\text{SNR} \rightarrow \infty} \{C(\text{SNR}) - \log \log \text{SNR}\} \quad (11)$$

and computed its value for different fading channels. For instance, when the fading is a zero-mean, unit-variance, circularly-symmetric, complex Gaussian process of spectral distribution function $F(\cdot)$, the fading number is [3, Cor. 4.42]

$$\chi = -1 - \gamma + \log \frac{1}{\epsilon^2} \quad (12)$$

where $\gamma \approx 0.577$ denotes Euler's constant, and where ϵ^2 is defined in (5).

It follows from (11) that, at high SNR, the capacity can be approximated as

$$C(\text{SNR}) \approx \log \log \text{SNR} + \chi. \quad (13)$$

Thus, at high SNR, communication is very power-inefficient, since one should expect to *square* the SNR for every additional bit per channel use. Since $\log \log \text{SNR}$ grows very slowly with the SNR, it follows that, over a large range of SNR, $\log \log \text{SNR}$ does not change much. For example, for $\text{SNR} \in [30\text{dB}, 80\text{dB}]$, it is between 2.1 and 3, and the capacity can be approximately bounded by

$$2.1 + \chi \leq C(\text{SNR}) \leq 3 + \chi, \quad \text{SNR} \in [30\text{dB}, 80\text{dB}]. \quad (14)$$

This gives rise to the rule of thumb that a system operating at rates considerably larger than $2 + \chi$ is probably operating in the high-SNR regime and is thus very power-inefficient [9], see also [10, 5]. The fading number can therefore be viewed as an indication of the maximal rate at which power-efficient communication is feasible.

While the fading number indicates at what *rates* communication is power-inefficient, it should be noted that it is difficult to determine the *SNR* at which this happens. Indeed, the fading number of zero-mean Gaussian fading channels depends on the spectral distribution function $F(\cdot)$ only via the mean-square error ϵ^2 in predicting the present fading from its past, whereas the SNR at which (13) becomes accurate depends not only on ϵ^2 , but also on $F(\cdot)$ itself [10, 5, 11].²

In the following section, we present results on the fading number of fading relay channels. They indicate at what rates a relay channel operates in the power-inefficient regime.

4 Main Results

Following Lapidoth and Moser [3], we define the fading number of the relay channel as

$$\chi \triangleq \overline{\lim}_{\text{SNR} \rightarrow \infty} \{C(\text{SNR}) - \log \log \text{SNR}\}. \quad (15)$$

An upper bound follows from the so called *max-flow min-cut upper bound* [12, Th. 14.7.1].

Theorem 1 (Upper bound). *Consider the above fading relay channel. Then*

$$\chi \leq \min \left\{ -2\gamma + \log \frac{1}{\epsilon_1^2} + \log \frac{1}{\epsilon_2^2}, \max \left\{ -1 - \gamma + \log \frac{1}{\epsilon_2^2}, -1 - \gamma + \log \frac{1}{\epsilon_3^2} \right\} \right\} \quad (16)$$

¹Note that (6) and (7) are fully characterized by SNR and ρ . Furthermore, we will see later that, in the limit as the SNR tends to infinity, the asymptotic behavior of $C(\text{SNR})$ does not depend on ρ . Thus, for a fixed ρ , the high-SNR asymptotic behavior of $C(\text{SNR})$ depends only on the SNR.

²The SNR at which (13) becomes accurate depends on the so-called *noisy prediction error* [5, Eq. (11)].

which becomes

$$\chi \leq -1 - \gamma + \log \frac{1}{\epsilon_2^2}, \quad \text{for } \epsilon_2^2 \leq \epsilon_3^2. \quad (17)$$

Here $\gamma \approx 0.577$ denotes Euler's constant, and ϵ_ℓ^2 , $\ell = 1, 2, 3$ are defined in (5).

Proof. See Section 6; equation (17) follows because $\epsilon_1^2 \leq 1$ and because $-2\gamma > -1 - \gamma$. \square

Note that $-1 - \gamma - \log \epsilon_2^2$ is the fading number of the fading channel between the transmitter and the receiver, whereas $-1 - \gamma - \log \epsilon_3^2$ is the fading number of the fading channel between the relay and the receiver (12). Thus, denoting the fading number of the former channel by χ_2 , and denoting the fading number of the latter channel by χ_3 , the upper bound (16) can be further upper-bounded by

$$\chi \leq \max\{\chi_2, \chi_3\}. \quad (18)$$

The right-hand side (RHS) of (18) is the fading number of a multiple-input single-output (MISO) fading channel with two transmit antennas and one receive antenna, where the fading processes corresponding to the different transmit antennas are independent, zero-mean, circularly-symmetric, complex Gaussian processes of spectral distribution function $F_\ell(\cdot)$, $\ell = 2, 3$ [13], see also [11, 14]. Thus, the fading number of the fading relay channel is upper-bounded by the fading number of the MISO channel that results when the transmitter and the relay can cooperate. In the following, we shall refer to this channel as the *TRC-MISO channel*. (Here “TRC” stands for “transmitter-relay cooperation”.)

It follows from (18) that if the fading number of the channel between the transmitter and the receiver is larger than the fading number of the channel between the relay and the receiver, i.e.,

$$\chi_2 \geq \chi_3$$

then at high SNR it is optimal to switch the relay off.

Corollary 2. *Let the fading processes $\{H_{2,k}, k \in \mathbb{Z}\}$ and $\{H_{3,k}, k \in \mathbb{Z}\}$ satisfy*

$$\epsilon_2^2 \leq \epsilon_3^2. \quad (19)$$

Then, the fading number of the above relay channel is given by

$$\chi = -1 - \gamma + \log \frac{1}{\epsilon_2^2}. \quad (20)$$

Using a *decode-and-forward* strategy [15], the following rates are achievable:

Theorem 3 (Lower bound). *Consider the above fading relay channel. Then*

$$\chi \geq \max\left\{-1 - \gamma + \log \frac{1}{\epsilon_2^2}, -1 - \gamma + \log \frac{1}{\epsilon_3^2} - \log\left(1 + \frac{\epsilon_1^2}{\epsilon_3^2}\right)\right\}. \quad (21)$$

Proof. See Section 7. \square

Denoting the fading number corresponding to the fading $\{H_{\ell,k}, k \in \mathbb{Z}\}$ by χ_ℓ , the lower bound (21) can be written as

$$\chi \geq \max\left\{\chi_2, \chi_3 - \log\left(1 + \exp(\chi_3 - \chi_1)\right)\right\} = \max\left\{\chi_2, \chi_1 - \log\left(1 + \exp(\chi_1 - \chi_3)\right)\right\}. \quad (22)$$

Note that if $\chi_1 \geq \chi_3$, then

$$\log\left(1 + \exp(\chi_3 - \chi_1)\right) \leq \log 2$$

and the difference between the lower bound (21) and the upper bound (18) is at most one bit.

Corollary 4. *Let the fading processes $\{H_{1,k}, k \in \mathbb{Z}\}$ and $\{H_{3,k}, k \in \mathbb{Z}\}$ satisfy*

$$\epsilon_1^2 \leq \epsilon_3^2. \quad (23)$$

Then, the fading number of the above fading relay channel is bounded by

$$\max \left\{ -1 - \gamma + \log \frac{1}{\epsilon_2^2}, -1 - \gamma + \log \frac{1}{\epsilon_3^2} \right\} \geq \chi \geq \max \left\{ -1 - \gamma + \log \frac{1}{\epsilon_2^2}, -1 - \gamma + \log \frac{1}{\epsilon_3^2} \right\} - \log 2. \quad (24)$$

As observed above, for SNRs below 80dB, the capacity is approximately upper-bounded by

$$C(\text{SNR}) \leq 3 + \chi, \quad \text{SNR} \leq 80\text{dB} \quad (25)$$

so a gap of $\log 2 \approx 0.6931$ nats seems substantial. Nevertheless, for slowly-varying fading channels, the prediction errors ϵ_ℓ^2 , $\ell = 1, 2, 3$ are small and the fading number, which depends on ϵ_ℓ^2 via $-\log \epsilon_\ell^2$, is much larger than $\log 2$. For example, for mobile speeds of the order of 5 km/h, prediction errors ϵ_ℓ^2 of roughly 10^{-4} seem plausible, see, e.g., [16, Sec. II]. In this case, the fading number is approximately

$$\chi = -1 - \gamma + \log \frac{1}{\epsilon_\ell^2} \approx 7.6331 \text{ nats} \quad (26)$$

and the RHS of (25) becomes 10.6331 nats. Thus, for slowly-varying fading channels, a gap of one bit (or equivalently $\log 2$ nats) is reasonably small.

Corollary 4 demonstrates that, when $\chi_1 \geq \chi_3$, the decode-and-forward scheme achieves communication rates that are within one bit of the capacity of the relay channel. This is consistent with the *Gaussian relay channel* where the decode-and-forward scheme achieves rates that are within one bit of the capacity, too [17, Th. 3.1]. Note that the difference between the lower bound (21) and the upper bound (18) decreases as $(\chi_1 - \chi_3)$ increases.

Thus, if the fading between the transmitter and the relay can be predicted more accurately than the fading between the relay and the receiver, then the fading number of the fading relay channel is at most one bit smaller than the fading number of the TRC-MISO channel. If we view the fading number as an indication of the rates at which communication is power-inefficient, then this result demonstrates that the rates at which the fading relay channel and the TRC-MISO channel operate in the power-inefficient regime are within one bit. Note however that this does not imply that for both channels the power-inefficient regime starts at the same SNR. Indeed, in the following section we derive nonasymptotic upper and lower bounds on the capacity of the fading relay channel as well as on the capacity of the TRC-MISO channel. These bounds suggest that the capacity of the fading relay channel increases much more slowly with the SNR than the capacity of the TRC-MISO channel.

5 Nonasymptotic Bounds

To simplify the analysis, we assume throughout this section that the channel between the transmitter and the receiver is memoryless, i.e.,

$$F_2'(\lambda) = 1, \quad -\frac{1}{2} \leq \lambda \leq \frac{1}{2}$$

which yields $\epsilon_2^2 = 1$. For this case, a nonasymptotic upper bound can be derived by extending [11, Eq. (16)] (see also [18, Th. 4.2]) to the above fading relay channel, i.e.,

$$C(\text{SNR}) \leq C_{\text{IID}}(\text{SNR}) + \log \left(1 + \frac{1}{\rho^2 \text{SNR}} \right) - \int_{1/2}^{1/2} \log \left(F_3'(\lambda) + \frac{1}{\rho^2 \text{SNR}} \right) d\lambda, \quad \text{SNR} > 0 \quad (27)$$

where $C_{\text{IID}}(\text{SNR})$ denotes the capacity in the memoryless fading case. It can be upper-bounded by (cf. [3, Eq. (141)])

$$C_{\text{IID}}(\text{SNR}) \leq \inf_{\substack{\alpha, \beta < 0, \\ \delta > 0}} \left\{ -1 + \alpha \log \frac{\beta}{\delta} + \log \Gamma \left(\alpha, \frac{\delta}{\beta} \right) + \log \delta \right. \\ \left. - (1 - \alpha) e^\delta \text{Ei}(-\delta) + \frac{\text{SNR}(1 + \rho^2) + 1}{\beta} + \frac{\delta}{\beta} \right\} \quad (28)$$

where

$$\Gamma(\nu, \xi) \triangleq \int_{\xi}^{\infty} t^{\nu-1} e^{-t} dt, \quad (\nu > 0, \xi \geq 0)$$

denotes the incomplete Gamma function, and where

$$\text{Ei}(-x) \triangleq - \int_x^{\infty} \frac{e^{-t}}{t} dt, \quad x > 0$$

denotes the exponential integral function. Note that the upper bound (27) is also an upper bound on the capacity of the TRC-MISO channel.

The next proposition presents a nonasymptotic lower bound on the capacity. It is similar to a lower bound that was derived in [18, Prop. 4.1] for single-antenna point-to-point fading channels with memory. For point-to-point fading channels, this bound is tight at high SNR in the sense that it achieves the fading number. For the fading relay channel, it can be shown that this bound achieves the lower bound on the fading number given in Theorem 3.

Proposition 5. *Let the fading process $\{H_{2,k}, k \in \mathbb{Z}\}$ be memoryless, i.e., let*

$$F_2'(\lambda) = 1, \quad -\frac{1}{2} \leq \lambda \leq \frac{1}{2}.$$

Then, we have

$$C(\text{SNR}) \geq \sup_{0 < \delta, \alpha, \delta_r < 1} \min\{R_{tr}(\text{SNR}; \delta, \alpha), R_{rr}(\text{SNR}, \rho; \delta, \alpha, \delta_r)\}, \quad (\text{SNR} > 0, \rho > 0) \quad (29)$$

where

$$\begin{aligned} R_{tr}(\text{SNR}; \delta, \alpha) &\triangleq \log\left(\frac{\sigma^{2(1-\alpha)}}{\delta^\alpha \text{SNR}^\alpha}\right) - \int_{-1/2}^{1/2} \log\left(F_1'(\lambda) + \frac{\sigma^{2(1-\alpha)}}{\delta^{2\alpha} \text{SNR}^\alpha}\right) d\lambda \\ &\quad - \exp\left(\frac{\sigma^{2(1-\alpha)} e}{\alpha \log(\frac{1}{\delta^2}) \delta^\alpha \text{SNR}^\alpha}\right) \text{Ei}\left(-\frac{\sigma^{2(1-\alpha)} e}{\alpha \log(\frac{1}{\delta^2}) \delta^\alpha \text{SNR}^\alpha}\right) \end{aligned} \quad (30)$$

and

$$\begin{aligned} R_{rr}(\text{SNR}, \rho; \delta, \alpha, \delta_r) &\triangleq \log \log \frac{1}{\delta_r^2} - 1 + \log\left(\frac{e^{-\gamma} \alpha \log(\frac{1}{\delta^2}) \delta^\alpha \text{SNR}^\alpha \sigma^{2(\alpha-1)} + e}{\log(\frac{1}{\delta_r^2}) \delta_r \rho^2 \text{SNR}}\right) \\ &\quad - \int_{-1/2}^{1/2} \log\left(F_3'(\lambda) + \frac{\sigma^{2(\alpha-1)}}{\delta_r^2 \rho^2 \text{SNR}^{1-\alpha}} + \frac{1}{\delta_r^2 \rho^2 \text{SNR}}\right) d\lambda \\ &\quad - \exp\left(\frac{e^{-\gamma} \alpha \log(\frac{1}{\delta^2}) \delta^\alpha \text{SNR}^\alpha \sigma^{2(\alpha-1)} + e}{\log(\frac{1}{\delta_r^2}) \delta_r \rho^2 \text{SNR}}\right) \text{Ei}\left(-\frac{e^{-\gamma} \alpha \log(\frac{1}{\delta^2}) \delta^\alpha \text{SNR}^\alpha \sigma^{2(\alpha-1)} + e}{\log(\frac{1}{\delta_r^2}) \delta_r \rho^2 \text{SNR}}\right). \end{aligned} \quad (31)$$

Proof. See Appendix A. □

A lower bound on the capacity of the TRC-MISO channel (but not necessarily the relay channel) follows by using a beam-selection strategy, where the transmitter transmits either from the first antenna (i.e., the transmitter) or from the second antenna (i.e., the relay). The TRC-MISO capacity is thus lower-bounded by [18, Prop. 4.1]

$$C_{\text{MISO}}(\text{SNR}) \geq \sup_{0 < \delta < 1} \max\{R_2(\text{SNR}, \rho; \delta), R_3(\text{SNR}, \rho; \delta)\}, \quad (\text{SNR} > 0, \rho > 0) \quad (32)$$

where

$$\begin{aligned} R_\ell(\text{SNR}, \rho; \delta) &\triangleq \log\left(\frac{1}{\delta \text{SNR}(1 + \rho^2)}\right) - \int_{-1/2}^{1/2} \log\left(F_\ell'(\lambda) + \frac{1}{\delta^2 \text{SNR}(1 + \rho^2)}\right) d\lambda \\ &\quad - \exp\left(\frac{e}{\log(\frac{1}{\delta^2}) \delta \text{SNR}(1 + \rho^2)}\right) \text{Ei}\left(-\frac{e}{\log(\frac{1}{\delta^2}) \delta \text{SNR}(1 + \rho^2)}\right), \quad \ell = 2, 3. \end{aligned} \quad (33)$$

Note that beam-selection is optimal at high SNR in the sense that it achieves the fading number [13, 11, 14].

While the above lower bounds (29) and (32) are tight at high SNR, they are loose at low SNR. We therefore include the following lower bounds that are superior to (29) and (32) when the SNR is small.

A lower bound on the capacity of the TRC-MISO channel follows by using a beam-selection strategy, and by lower-bounding the capacity of the resulting point-to-point channel by choosing quaternary phase-shift keying (QPSK) channel inputs [19, Prop. 2.1 & Eq. (17)], i.e.,

$$C_{\text{MISO}}(\text{SNR}) \geq \max \left\{ I(X; (\bar{H}_1 + \tilde{H}_1)X + Z_{r,1} \mid \bar{H}_1), I(X_r; (\bar{H}_3 + \tilde{H}_3)X_r + Z_1 \mid \bar{H}_3) \right\} \quad (34)$$

where \bar{H}_ℓ , $\ell = 1, 3$ and \tilde{H}_ℓ , $\ell = 1, 3$ are independent, zero-mean, circularly-symmetric, complex Gaussian random variables of variance $1 - \epsilon_\ell^2((1 + \rho^2)\mathcal{A}^2)$ and $\epsilon_\ell^2((1 + \rho^2)\mathcal{A}^2)$, respectively; and X and X_r are uniformly distributed over

$$\left\{ \sqrt{(1 + \rho^2)}\mathcal{A}, i\sqrt{(1 + \rho^2)}\mathcal{A}, -\sqrt{(1 + \rho^2)}\mathcal{A}, -i\sqrt{(1 + \rho^2)}\mathcal{A} \right\}.$$

Here the prediction errors are

$$\epsilon_\ell(\xi) \triangleq \exp \left(\int_{-1/2}^{1/2} \log \left(F'_\ell(\lambda) - \frac{1}{\xi} \right) d\lambda \right) - \frac{1}{\xi}, \quad \ell = 1, 3. \quad (35)$$

The RHS of (34) can be computed numerically.

The above lower bound can be extended to the fading relay channel by employing a decode-and-forward strategy (Proposition 6), and by choosing $\{X_k, k \in \mathbb{Z}\}$ and $\{X_{r,k}, k \in \mathbb{Z}\}$ to be i.i.d. random variables, independent of each other and with

- X being uniformly distributed over $\{\delta\mathcal{A}, i\delta\mathcal{A}, -\delta\mathcal{A}, -i\delta\mathcal{A}\}$, for some $0 < \delta < 1$;
- X_r being uniformly distributed over $\{\mathcal{A}, i\mathcal{A}, -\mathcal{A}, -i\mathcal{A}\}$.

This yields

$$C(\text{SNR}) \geq \sup_{0 < \delta < 1} \min \left\{ I(X; (\bar{H}'_1 + \tilde{H}'_1)X + Z_{r,1} \mid \bar{H}'_1), I(X_r; (\bar{H}_3 + \tilde{H}_3)X_r + H_{2,1}X + Z_1 \mid \bar{H}_3, X_1) \right\} \quad (36)$$

where \bar{H}_3 and \tilde{H}_3 are as above, and \bar{H}'_1 and \tilde{H}'_1 are independent, zero-mean, circularly-symmetric, complex Gaussian random variables of variance $1 - \epsilon_1^2(\delta^2\mathcal{A}^2)$ and $\epsilon_1^2(\delta^2\mathcal{A}^2)$, respectively. The RHS of (36) can be computed numerically.

We evaluate the upper bound (27) and the lower bounds (29), (32), (34), and (36) for spectral distribution functions of the form

$$F'_\ell(\lambda) = \begin{cases} \Upsilon_\ell, & |\lambda| \leq \Delta_\ell \\ \Lambda_\ell, & \Delta_\ell < |\lambda| \leq \frac{1}{2}, \end{cases} \quad \ell = 1, 3 \quad (37)$$

where $\Upsilon_\ell > 0$, $\Lambda_\ell > 0$, and $0 < \Delta_\ell < 1/2$ satisfy

$$\int_{-1/2}^{1/2} F'_\ell(\lambda) d\lambda = 2\Upsilon_\ell\Delta_\ell + (1 - 2\Delta_\ell)\Lambda_\ell = 1, \quad \ell = 1, 3. \quad (38)$$

(Recall that $F'_2(\lambda) = 1$, $1/2 \leq \lambda \leq 1/2$.) We shall consider two scenarios: in the first scenario, the fading between the transmitter and the relay has a prediction error of 10^{-4} , whereas the fading between the relay and the receiver has a prediction error of 10^{-2} . This implies that the fading number of the relay channel is roughly the same as the fading number of the TRC-MISO channel. In the second scenario, both the fading between the transmitter and the relay and the fading between the relay and the receiver have a prediction error of 10^{-2} . In this case, the lower bound

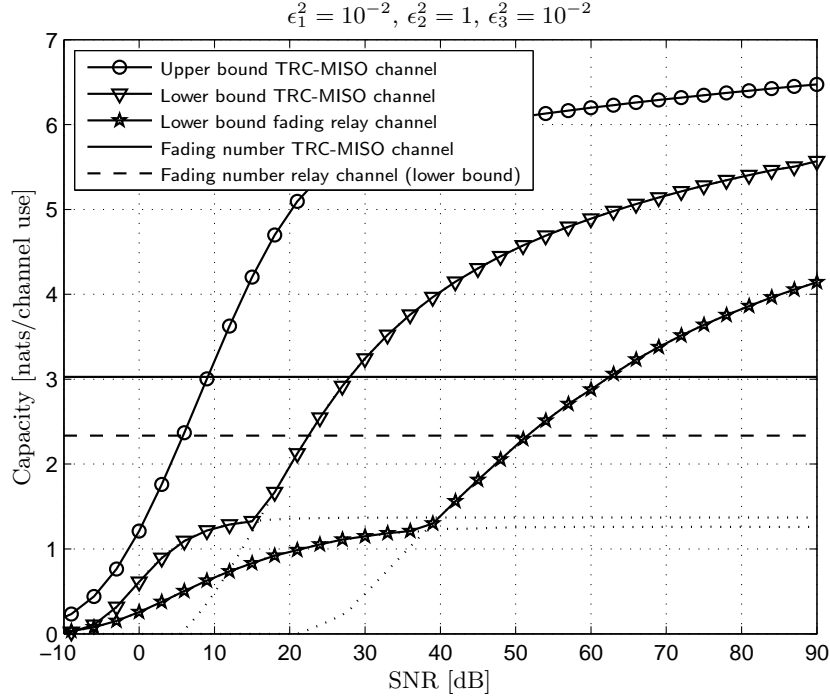
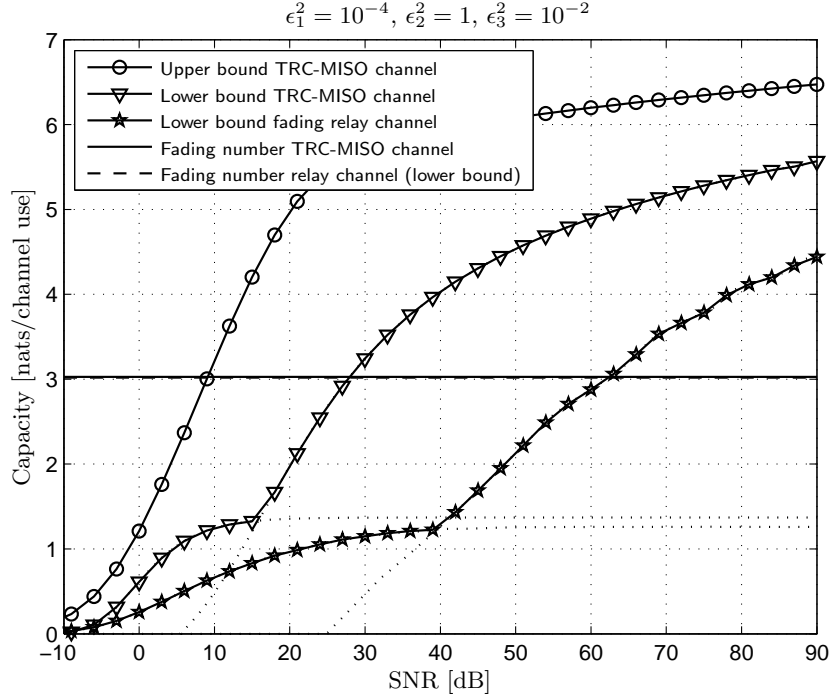


Figure 2: Upper bound on the capacity of the TRC-MISO channel (27); lower bounds on the capacity of the TRC-MISO channel [maximum of (32) and (34)] and on the capacity of the fading relay channel [maximum of (29) and (36)]; the fading number of the TRC-MISO channel and the lower bound (21) on the fading number of the fading relay channel. The prediction errors are $\epsilon_1^2 = 10^{-4}, \epsilon_2^2 = 1, \epsilon_3^2 = 10^{-2}$ (top figure) and $\epsilon_1^2 = 10^{-2}, \epsilon_2^2 = 1, \epsilon_3^2 = 10^{-2}$ (bottom figure).

on the fading number of the relay channel (21) is $\log 2$ nats smaller than the fading number of the TRC-MISO channel.

Figure 2 shows the upper bound on the capacity of the fading relay channel and of the TRC-MISO channel (27), the lower bounds on the capacity of the fading relay channel [maximum of (29) and (36)] and of the TRC-MISO channel [maximum of (32) and (34)], together with the corresponding fading numbers for the above two scenarios. In particular, the top figure in Figure 2 shows the bounds (27), (29), (32), (34), and (36) for

$$\begin{array}{llll} \Upsilon_1 & \approx & 5.76034 & \quad \quad \quad \Upsilon_3 \approx 10.99684 \\ \Lambda_1 & = & 10^{-5} & \quad \quad \quad \Lambda_3 = 0.005 \\ \Delta_1 & \approx & 0.08679 & \quad \quad \quad \Delta_3 \approx 0.04503 \end{array} \quad \text{and}$$

resulting in $\epsilon_1^2 = 10^{-4}$ and $\epsilon_3^2 = 10^{-2}$. In this case, the fading number is upper-bounded by (16)

$$\chi \leq -1 - \gamma + \log \frac{1}{\epsilon_3^2} \approx 3.0280 \quad (39)$$

which is equal to the fading number of the TRC-MISO channel. The fading number of the relay channel is lower-bounded by (21)

$$\chi \geq -1 - \gamma + \log \frac{1}{\epsilon_3^2} - \log \left(1 + \frac{\epsilon_1^2}{\epsilon_3^2} \right) \approx 3.0180. \quad (40)$$

The bottom figure in Figure 2 shows the bounds (27), (29), (32), (34), and (36) for

$$\begin{array}{llll} \Upsilon_1 & = & \Upsilon_3 & \approx 10.99684 \\ \Lambda_1 & = & \Lambda_3 & = 0.005 \\ \Delta_1 & = & \Delta_3 & \approx 0.04503 \end{array}$$

resulting in $\epsilon_1^2 = \epsilon_3^2 = 10^{-2}$. As in the above example, for the relay channel this yields

$$\chi \leq -1 - \gamma + \log \frac{1}{\epsilon_3^2} \approx 3.0280 \quad (41)$$

which is equal to the fading number of the TRC-MISO fading channel. The lower bound (21) becomes

$$\chi \geq -1 - \gamma + \log \frac{1}{\epsilon_3^2} - \log 2 \approx 2.3348. \quad (42)$$

In both cases we assume that $\rho = 1$ and $\sigma = 1$.

Note that for the above spectral distribution functions the fading processes $\{H_{1,k}, k \in \mathbb{Z}\}$ and $\{H_{3,k}, k \in \mathbb{Z}\}$ are *nonpermal* in the sense that [20, Def. 2.1]

$$\int_{-1/2}^{1/2} F_\ell'^2(\lambda) d\lambda > 2, \quad \ell = 1, 3. \quad (43)$$

In this case i.i.d. inputs and QPSK as well as beam-selection achieve the low-SNR asymptotic capacity [20, Secs. II-A3 & II-B4]. Thus, for the above spectral distribution functions, the lower bound (34) is tight at low SNR.

We observe that the lower bound for the fading relay channel (29) increases much more slowly with the SNR than the lower bound for the TRC-MISO channel (32), even in the first example where the fading numbers of both channels are almost identical. Since these lower bounds are tight at high SNR, we suspect that the same is also true for the capacities of both channels at high SNR. Thus, even though the capacities of the fading relay channel and the TRC-MISO channel have similar asymptotic behaviors in the limit as the SNR tends to infinity, they may differ substantially at finite SNR.

6 Proof of Upper Bound

Theorem 1 follows from Fano's inequality [12, Th. 2.11.1] and from the following upper bound on $\frac{1}{n}I(X_1^n; Y_1^n)$

$$\begin{aligned} C(\text{SNR}) &\leq \lim_{n \rightarrow \infty} \sup \frac{1}{n} I(X_1^n; Y_1^n) \\ &\leq \min \left\{ \lim_{n \rightarrow \infty} \sup \frac{1}{n} I(X_1^n; Y_{r,1}^n, Y_1^n), \lim_{n \rightarrow \infty} \frac{1}{n} \sup I(X_1^n, X_{r,1}^n; Y_1^n) \right\} \end{aligned} \quad (44)$$

where the suprema are over all joint distributions on $(X_1^n, X_{r,1}^n)$ satisfying the power constraints (6) and (7). Here the second step follows by upper-bounding

$$I(X_1^n; Y_1^n) \leq I(X_1^n; Y_{r,1}^n, Y_1^n) \quad \text{and} \quad I(X_1^n; Y_1^n) \leq I(X_1^n, X_{r,1}^n; Y_1^n)$$

which in turn follows because $I(A; B) \leq I(A; B, C)$ for every random variables A , B , and C .

The first term on the RHS of (44) is upper-bounded by the capacity of a single-input multiple-output (SIMO) fading channel with peak-power A^2 , whereas the second term on the RHS of (44) is upper-bounded by the capacity of a MISO fading channel with peak-power $A^2 + A_r^2$, which by (8) is equal to $A^2(1 + \rho^2)$. Indeed, we can upper-bound the first term on the RHS of (44) as follows:

$$\begin{aligned} &\frac{1}{n} I(X_1^n; Y_{r,1}^n, Y_1^n) \\ &\leq \frac{1}{n} I(X_1^n; Y_{r,1}^n, Y_1^n \mid H_{3,1}^n) \\ &= \frac{1}{n} \sum_{k=1}^n I(X_1^n; Y_{r,k}, Y_k \mid H_{3,1}^n, Y_{r,1}^{k-1}, Y_1^{k-1}) \\ &= \frac{1}{n} \sum_{k=1}^n I(X_1^k; Y_{r,k}, Y_k \mid H_{3,1}^k = 0, Y_{r,1}^{k-1}, Y_1^{k-1}) \\ &\leq \frac{1}{n} \sum_{k=1}^n I(X_1^k, Y_{r,1}^{k-1}, Y_1^{k-1}; Y_{r,k}, Y_k \mid H_{3,1}^k = 0) \\ &\leq \frac{1}{n} \sum_{k=1}^n I(X_k, H_{1,1}^{k-1}, H_{2,1}^{k-1}; Y_{r,k}, Y_k \mid H_{3,1}^k = 0) \\ &\leq \frac{1}{n} \sum_{k=1}^n \sup I(X_k; Y_{r,k}, Y_k \mid H_{3,k} = 0) + \frac{1}{n} \sum_{k=1}^n I(H_{1,1}^{k-1}, H_{2,1}^{k-1}; Y_{r,k}, Y_k \mid X_k, H_{3,k} = 0) \\ &\leq \sup I(X_1; Y_{r,1}, Y_1 \mid H_{3,1} = 0) + \frac{1}{n} \sum_{k=1}^n I(H_{1,1}^{k-1}, H_{2,1}^{k-1}; Y_{r,k}, Y_k \mid X_k, H_{3,k} = 0) \end{aligned} \quad (45)$$

where the first supremum is over all input distributions on X_k satisfying (6), and the second supremum is over all input distributions on X_1 satisfying (6). Here the first step follows because X_1^n is independent of $H_{3,1}^n$; the second step follows from the chain rule for mutual information [12, Th. 2.5.2]; the third step follows because $X_{r,1}^k$ is a function of $Y_{r,1}^{k-1}$, so $(H_{3,1}^n, X_{r,1}^k)$ is known and we can therefore subtract $H_{3,\ell} X_{r,\ell}$ from Y_ℓ , $\ell = 1, \dots, k$, resulting in the same mutual information as if we would set $H_{3,1}^k = 0$; the fourth step follows because $I(A; B|C) \leq I(A, C; B)$ for any random variables A , B , and C (this is a consequence of the chain rule for mutual information and of the non-negativity of mutual information); the fifth step follows by adding the observations $(H_{1,1}^{k-1}, H_{2,1}^{k-1})$, and by noting that, conditional on $(X_k, H_{3,1}^k = 0, H_{1,1}^{k-1}, H_{2,1}^{k-1})$, the pair $(Y_{r,k}, Y_k)$ is independent of $(X_1^{k-1}, Y_{r,1}^{k-1}, Y_1^{k-1})$; the sixth step follows from the chain rule for mutual information and by upper-bounding each summand in the first sum by its supremum; and the last step follows from the stationarity of the channel, which implies that $\sup I(X_k; Y_{r,k}, Y_k \mid H_{3,k} = 0)$ does not depend on k .

The first term on the RHS of (45) is the capacity of a memoryless SIMO fading channel, which is given by [3, Cor. 4.32]

$$\sup I(X_1; Y_{r,1}, Y_1 \mid H_{3,1} = 0) = \log \log \text{SNR} - 2\gamma + o(1) \quad (46)$$

where $o(1)$ tends to zero as $\text{SNR} \rightarrow \infty$. The second term on the RHS of (45) can be upper-bounded by

$$\begin{aligned} \frac{1}{n} \sum_{k=1}^n I(H_{1,1}^{k-1}, H_{2,1}^{k-1}; Y_{r,k}, Y_k \mid X_k, H_{3,k} = 0) &\leq \frac{1}{n} \sum_{k=1}^n I(H_{1,1}^{k-1}, H_{2,1}^{k-1}; H_{1,k}, H_{2,k}) \\ &= \frac{1}{n} \sum_{k=1}^n \left[I(H_{1,1}^{k-1}; H_{1,k}) + I(H_{2,1}^{k-1}; H_{2,k}) \right] \end{aligned} \quad (47)$$

where the first step follows by adding the observations $(H_{1,k}, H_{2,k})$ and by noting that, conditioned on $(X_k, H_{3,k} = 0, H_{1,k}, H_{2,k})$, the pair $(Y_{r,k}, Y_k)$ is independent of $(H_{1,1}^{k-1}, H_{2,1}^{k-1})$; and the second step follows because the processes $\{H_{1,k}, k \in \mathbb{Z}\}$ and $\{H_{2,k}, k \in \mathbb{Z}\}$ are independent.

By Cesàro's mean [12, Th. 4.2.3], it follows that

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \left[I(H_{1,1}^{k-1}; H_{1,k}) + I(H_{2,1}^{k-1}; H_{2,k}) \right] &= \lim_{k \rightarrow \infty} I(H_{1,1}^{k-1}; H_{1,k}) + \lim_{k \rightarrow \infty} I(H_{2,1}^{k-1}; H_{2,k}) \\ &= \log \frac{1}{\epsilon_1^2} + \log \frac{1}{\epsilon_2^2}. \end{aligned} \quad (48)$$

Here the second step follows because $\{H_{1,k}, k \in \mathbb{Z}\}$ and $\{H_{2,k}, k \in \mathbb{Z}\}$ are unit-variance Gaussian processes whose conditional variances, conditioned on the past $(k-1)$ fading coefficients, tend to ϵ_1^2 and ϵ_2^2 as k tends to infinity [21, Lemmas 5.7(b) & 5.10(c)]. Combining (45)–(48), we thus obtain

$$\lim_{n \rightarrow \infty} \sup \frac{1}{n} I(X_1^n; Y_{r,1}^n, Y_1^n) \leq \log \log \text{SNR} - 2\gamma + \log \frac{1}{\epsilon_1^2} + \log \frac{1}{\epsilon_2^2} + o(1). \quad (49)$$

To evaluate the second term on the RHS of (44), we first note that, by (6), (7), and (8), the channel inputs X_k and $X_{r,k}$ satisfy

$$|X_k|^2 + |X_{r,k}|^2 \leq A^2 (1 + \rho^2), \quad k \in \mathbb{Z} \quad (50)$$

with probability one. By maximizing over all joint distributions on $(X_1^n, X_{r,1}^n)$ satisfying (50) [rather than (6) and (7)], it follows that $\lim_{n \rightarrow \infty} \sup \frac{1}{n} I(X_1^n, X_{r,1}^n; Y_1^n)$ is upper-bounded by the capacity of a MISO fading channel with fading processes $\{H_{2,k}, k \in \mathbb{Z}\}$ and $\{H_{3,k}, k \in \mathbb{Z}\}$ and with peak-power constraint $A^2 (1 + \rho^2)$, i.e., we have [13, Cor. 8] (see also [11, Cor. 8], [14, Cor. 5.6]),

$$\lim_{n \rightarrow \infty} \sup \frac{1}{n} I(X_1^n, X_{r,1}^n; Y_1^n) \leq \log \log (\text{SNR} (1 + \rho^2)) - 1 - \gamma + \max \left\{ \log \frac{1}{\epsilon_2^2}, \log \frac{1}{\epsilon_3^2} \right\} + o(1). \quad (51)$$

Combining (49) and (51) with (44), we obtain

$$\begin{aligned} C(\text{SNR}) &\leq \min \left\{ \log \log \text{SNR} - 2\gamma + \log \frac{1}{\epsilon_1^2} + \log \frac{1}{\epsilon_2^2} + o(1), \right. \\ &\quad \left. \log \log (\text{SNR} (1 + \rho^2)) - 1 - \gamma + \max \left\{ \log \frac{1}{\epsilon_2^2}, \log \frac{1}{\epsilon_3^2} \right\} + o(1) \right\}. \end{aligned} \quad (52)$$

Computing the difference $(C(\text{SNR}) - \log \log \text{SNR})$ in the limit as the SNR tends to infinity, and noting that

$$\lim_{\text{SNR} \rightarrow \infty} \{ \log \log (\text{SNR} (1 + \rho^2)) - \log \log \text{SNR} \} = 0$$

it follows that

$$\chi \leq \min \left\{ -2\gamma + \log \frac{1}{\epsilon_1^2} + \log \frac{1}{\epsilon_2^2}, \max \left\{ -1 - \gamma + \log \frac{1}{\epsilon_2^2}, -1 - \gamma + \log \frac{1}{\epsilon_3^2} \right\} \right\}. \quad (53)$$

This proves Theorem 1.

7 Proof of Lower Bound

In the following we prove Theorem 3. The first term in (21) is the fading number of the channel between the transmitter and receiver [3, Cor. 4.42] and follows by switching the relay off. In the following, we prove the remaining terms. They are based on the following proposition.

Proposition 6. *Consider the above fading relay channel. Then the rate*

$$R = \lim_{n \rightarrow \infty} \sup \frac{1}{n} \min \left\{ I(X_1^n; Y_{r,1}^n \mid X_{r,1}^n), I(X_1^n, X_{r,1}^n; Y_1^n) \right\} \quad (54)$$

is achievable. Here the supremum is over all product distributions on $(X_1^n, X_{r,1}^n)$, i.e.,

$$P_{X_1^n, X_{r,1}^n}(\cdot) = \prod_{k=1}^n P_{X, X_r}(\cdot)$$

satisfying the power constraints (6) and (7).

Proof. See Appendix B. □

Note that (54) is an extension of the *decode-and-forward* (DF) scheme [15, Th. 1] to channels with memory.

Theorem 3 follows from Proposition 6 upon choosing $\{X_k, k \in \mathbb{Z}\}$ and $\{X_{r,k}, k \in \mathbb{Z}\}$ to be i.i.d., circularly-symmetric random variables, independent of each other and with

$$\log |X_k|^2 \sim \mathcal{U}([\log \log A^2, \log A^{2\alpha}]), \quad k \in \mathbb{Z} \quad (55)$$

$$\log |X_{r,k}|^2 \sim \mathcal{U}([\log A_r^{2\beta}, \log A_r^2]), \quad k \in \mathbb{Z} \quad (56)$$

for some $0 < \alpha < \beta < 1$. Here $\mathcal{U}(\mathcal{A})$ denotes the uniform distribution over the set \mathcal{A} .

Before we set out to prove Theorem 3, we pause for intuition. Recall that if the channel between the transmitter and the receiver has a larger fading number than the channel between the relay and the receiver, then it is optimal to switch the relay off, i.e., to set $X_{r,k} = 0, k \in \mathbb{Z}$. This happens if $\chi_2 \geq \chi_3$, and we therefore focus on the case where $\chi_3 > \chi_2$. Since every signal sent from the transmitter to the relay interferes also at the receiver, there is a trade-off between achieving high data rates from the transmitter to the relay (requiring a large transmit-power) and minimizing the interference at the receiver (requiring a low transmit-power). We address this problem by choosing $P_{X, X_r}(\cdot)$ such that $\frac{X_k}{X_{r,k}}$ vanishes as A^2 tends to infinity, i.e., by allowing the relay a much larger transmit-power than the transmitter.

The input distribution (55) and (56) chooses X_k and $X_{r,k}$ independently and trades rates from the transmitter to the relay against rates from the relay to the receiver by using the parameters α and β . For instance, increasing α allows for larger rates between the transmitter and the relay, but requires a larger β (since we have $\beta > \alpha$), decreasing the rates achievable between the relay and the receiver.

7.1 Lower Bound on $\lim_{n \rightarrow \infty} \frac{1}{n} I(X_1^n; Y_{r,1}^n \mid X_{r,1}^n)$

We first lower-bound the first term on the RHS of (54). We have

$$\begin{aligned} I(X_1^n; Y_{r,1}^n \mid X_{r,1}^n) &= I(X_1^n; Y_{r,1}^n) \\ &= \sum_{k=1}^n I(X_k; Y_{r,1}^n \mid X_1^{k-1}) \\ &\geq \sum_{k=\kappa+1}^n I(X_k; Y_{r,1}^n \mid X_1^{k-1}) \end{aligned} \quad (57)$$

for every $0 \leq \kappa < n$. Here the first step follows because X_1^n and $X_{r,1}^n$ are independent, and because X_1^n and $X_{r,1}^n$ are also independent when conditioned on $Y_{r,1}^n$; the second step follows from the

chain rule for mutual information; and the third step follows from the nonnegativity of mutual information. Using that $\{X_k, k \in \mathbb{Z}\}$ is i.i.d. and that reducing observations does not increase mutual information, we obtain

$$\begin{aligned}
I(X_k; Y_{r,1}^n \mid X_1^{k-1}) &\geq I(X_k; Y_{r,1}^k \mid X_1^{k-1}) \\
&= I(X_k; Y_{r,1}^k, X_1^{k-1}) \\
&\geq I(X_k; Y_{r,k-\kappa}^k, X_{k-\kappa}^{k-1}) \\
&= I(X_k; Y_{r,k-\kappa}^k, H_{1,k-\kappa}^{k-1}, X_{k-\kappa}^{k-1}) - \varepsilon_1(\text{SNR}, \kappa)
\end{aligned} \tag{58}$$

where ε_1 is defined as

$$\begin{aligned}
\varepsilon_1(\text{SNR}, \kappa) &\triangleq I(X_k; Y_{r,k-\kappa}^k, H_{1,k-\kappa}^{k-1}, X_{k-\kappa}^{k-1}) - I(X_k; Y_{r,k-\kappa}^k, X_{k-\kappa}^{k-1}) \\
&= I(X_k; H_{1,k-\kappa}^{k-1} \mid Y_{r,k-\kappa}^k, X_{k-\kappa}^{k-1}).
\end{aligned}$$

Note that, due to the stationarity of the channel and of the proposed coding scheme, $\varepsilon_1(\text{SNR}, \kappa)$ does not depend on k . Furthermore, it follows from [3, App. IX] that for every fixed κ

$$\lim_{\text{SNR} \rightarrow \infty} \varepsilon_1(\text{SNR}, \kappa) = 0. \tag{59}$$

We further lower-bound the RHS of (58) by

$$\begin{aligned}
I(X_k; Y_{r,k-\kappa}^k, H_{1,k-\kappa}^{k-1}, X_{k-\kappa}^{k-1}) &\geq I(X_k; Y_{r,k}, H_{1,k-\kappa}^{k-1}) \\
&= I(X_k; Y_{r,k} \mid H_{1,k-\kappa}^{k-1})
\end{aligned} \tag{60}$$

which follows because reducing observations does not increase mutual information, and because $I(A; B, C) \geq I(A; B|C)$ for any random variables A, B , and C .

We express the fading coefficient $H_{1,k}$ at time k as

$$H_{1,k} = \bar{H}_{1,k} + \tilde{H}_{1,k}$$

where $\bar{H}_{1,k} = \mathbb{E}[H_{1,k} \mid H_{1,k-\kappa}^{k-1}]$ is the best predictor of $H_{1,k}$ given $H_{1,k-1}, \dots, H_{1,k-\kappa}$, and where $\tilde{H}_{1,k}$ denotes the prediction error. Note that, since $\{H_{1,k} \mid k \in \mathbb{Z}\}$ is a zero-mean, complex Gaussian process, it follows that also $\bar{H}_{1,k}$ and $\tilde{H}_{1,k}$ are zero-mean, complex Gaussian random variables with variance $1 - \epsilon_{1,\kappa}^2$ and $\epsilon_{1,\kappa}^2$, respectively. Further note that $\tilde{H}_{1,k}$ is independent of $H_{1,k-\kappa}^{k-1}$ [21, Lemma 5.8], and that [21, Lemmas 5.7(b) & 5.10(c)], [8]

$$\lim_{\kappa \rightarrow \infty} \epsilon_{1,\kappa}^2 = \epsilon_1^2 = \exp\left(\int_{-1/2}^{1/2} \log F'_1(\lambda) d\lambda\right). \tag{61}$$

With this, we obtain

$$\begin{aligned}
&I(X_k; Y_{r,k} \mid H_{1,k-\kappa}^{k-1}) \\
&= h\left((\bar{H}_{1,k} + \tilde{H}_{1,k})X_k + Z_{r,k} \mid \bar{H}_{1,k}\right) - h\left((\bar{H}_{1,k} + \tilde{H}_{1,k})X_k + Z_{r,k} \mid \bar{H}_{1,k}, X_k\right) \\
&\geq h(H_{1,k}X_k + Z_{r,k} \mid H_{1,k}, Z_{r,k}) - h\left((\bar{H}_{1,k} + \tilde{H}_{1,k})X_k + Z_{r,k} \mid \bar{H}_{1,k}, X_k\right) \\
&= h(H_{1,k}X_k \mid H_{1,k}) - h(\tilde{H}_{1,k}X_k + Z_{r,k} \mid X_k) \\
&= \mathbb{E}[\log |H_{1,k}|^2] + h(X_k) - \mathbb{E}[\log |X_k|^2] - h\left(\tilde{H}_{1,k} + \frac{Z_{r,k}}{X_k} \mid X_k\right) \\
&= \mathbb{E}[\log |H_{1,k}|^2] + h(X_k) - \mathbb{E}[\log |X_k|^2] - \log \pi - 1 - \mathbb{E}\left[\log\left(\epsilon_{1,\kappa}^2 + \frac{\sigma^2}{|X_k|^2}\right)\right] \\
&\geq \mathbb{E}[\log |H_{1,k}|^2] + h(X_k) - \mathbb{E}[\log |X_k|^2] - \log \pi - 1 - \log\left(\epsilon_{1,\kappa}^2 + \frac{\sigma^2}{\log A^2}\right)
\end{aligned} \tag{62}$$

where the second step follows because conditioning does not increase entropy; the third step follows from the behavior of differential entropy under translation [12, Th. 9.6.3]; the fourth step follows

from the behavior of differential entropy under scaling by a complex number [12, Th. 9.6.4]; the fifth step follows because, conditioned on X_k , the random variable $\tilde{H}_{1,k} + Z_{r,k}/X_k$ is Gaussian; and the last step follows because, for our choice of input distribution (55), we have $|X_k|^2 \geq \log A^2$ with probability one.

The first term on the RHS of (62) can be evaluated as [22, Sec. 4.331]

$$\mathbb{E}[\log |H_{1,k}|^2] = -\gamma. \quad (63)$$

The subsequent three terms yield [3, Lemmas 6.15 & 6.16]

$$h(X_k) - \mathbb{E}[\log |X_k|^2] - \log \pi = h(\log |X_k|^2) = \log(\alpha \log A^2 - \log \log A^2). \quad (64)$$

Combining (58)–(64), and noting that the RHS of (62) does not depend on k , yields

$$\begin{aligned} & \frac{1}{n} I(X_1^n; Y_{r,1}^n \mid X_{r,1}^n) \\ & \geq \frac{n-\kappa}{n} \left[\log(\alpha \log A^2 - \log \log A^2) - 1 - \gamma - \log \left(\epsilon_{1,\kappa}^2 + \frac{\sigma^2}{\log A^2} \right) - \varepsilon_1(\text{SNR}, \kappa) \right] \end{aligned} \quad (65)$$

which tends to

$$\log(\alpha \log A^2 - \log \log A^2) - 1 - \gamma - \log \left(\epsilon_{1,\kappa}^2 + \frac{\sigma^2}{\log A^2} \right) - \varepsilon_1(\text{SNR}, \kappa)$$

as n tends to infinity. Using (59), and noting that

$$\begin{aligned} \lim_{\text{SNR} \rightarrow \infty} \log(\alpha \log A^2 - \log \log A^2) - \log \log \text{SNR} &= \log \alpha \\ \lim_{\text{SNR} \rightarrow \infty} \log \left(\epsilon_{1,\kappa}^2 + \frac{\sigma^2}{\log A^2} \right) &= \log \epsilon_{1,\kappa}^2 \end{aligned}$$

we obtain

$$\lim_{\text{SNR} \rightarrow \infty} \left\{ \lim_{n \rightarrow \infty} \frac{1}{n} I(X_1^n; Y_{r,1}^n \mid X_{r,1}^n) - \log \log \text{SNR} \right\} \geq -1 - \gamma + \log \frac{1}{\epsilon_{1,\kappa}^2} + \log \alpha \quad (66)$$

which tends to

$$\lim_{\text{SNR} \rightarrow \infty} \left\{ \lim_{n \rightarrow \infty} \frac{1}{n} I(X_1^n; Y_{r,1}^n \mid X_{r,1}^n) - \log \log \text{SNR} \right\} \geq -1 - \gamma + \log \frac{1}{\epsilon_1^2} + \log \alpha \quad (67)$$

as κ tends to infinity.

7.2 Lower Bound on $\lim_{n \rightarrow \infty} \frac{1}{n} I(X_1^n, X_{r,1}^n; Y_1^n)$

We continue by lower-bounding the second term on the RHS of (54). The proof is similar to the proof of (67), and we will therefore skip some of the details. We start with the chain rule for mutual information to obtain

$$\begin{aligned} I(X_1^n, X_{r,1}^n; Y_1^n) &= \sum_{k=1}^n I(X_k, X_{r,k}; Y_1^n \mid X_1^{k-1}, X_{r,1}^{k-1}) \\ &= \sum_{k=1}^n I(X_k, X_{r,k}; Y_1^n, X_1^{k-1}, X_{r,1}^{k-1}) \\ &\geq \sum_{k=\kappa+1}^n I(X_k, X_{r,k}; Y_{k-\kappa}^k, X_{k-\kappa}^{k-1}, X_{r,k-\kappa}^{k-1}) \end{aligned} \quad (68)$$

for every $0 \leq \kappa < n$. Here the second step follows because $\{(X_k, X_{r,k}, k \in \mathbb{Z})\}$ is i.i.d. We next define $\varepsilon_2(\text{SNR}, \kappa)$ as

$$\begin{aligned} \varepsilon_2(\text{SNR}, \kappa) &\triangleq I(X_k, X_{r,k}; Y_{k-\kappa}^k, X_{k-\kappa}^{k-1}, X_{r,k-\kappa}^{k-1}, H_{3,k-\kappa}^{k-1}) - I(X_k, X_{r,k}; Y_{k-\kappa}^k, X_{k-\kappa}^{k-1}, X_{r,k-\kappa}^{k-1}) \\ &= I(X_k, X_{r,k}; H_{3,k-\kappa}^{k-1} \mid Y_{k-\kappa}^k, X_{k-\kappa}^{k-1}, X_{r,k-\kappa}^{k-1}). \end{aligned} \quad (69)$$

We show in Appendix C that for every fixed κ

$$\lim_{\text{SNR} \rightarrow \infty} \varepsilon_2(\text{SNR}, \kappa) = 0. \quad (70)$$

With this definition, every summand on the RHS of (68) can be lower-bounded by

$$\begin{aligned} I(X_k, X_{r,k}; Y_{k-\kappa}^k, X_{k-\kappa}^{k-1}, X_{r,k-\kappa}^{k-1}) &= I(X_k, X_{r,k}; Y_{k-\kappa}^k, X_{k-\kappa}^{k-1}, X_{r,k-\kappa}^{k-1}, H_{3,k-\kappa}^{k-1}) - \varepsilon_2(\text{SNR}, \kappa) \\ &= I(X_k, X_{r,k}; Y_{k-\kappa}^k, X_{k-\kappa}^{k-1}, X_{r,k-\kappa}^{k-1} \mid H_{3,k-\kappa}^{k-1}) - \varepsilon_2(\text{SNR}, \kappa) \\ &\geq I(X_k, X_{r,k}; Y_k \mid H_{3,k}^{k-1}) - \varepsilon_2(\text{SNR}, \kappa) \end{aligned} \quad (71)$$

where the second step follows because $\{H_{3,k}, k \in \mathbb{Z}\}$ is independent of $(X_k, X_{r,k})$; and the third step follows because reducing observations does not increase mutual information.

As above, we express the fading $H_{3,k}$ as

$$H_{3,k} = \bar{H}_{3,k} + \tilde{H}_{r,k}$$

where $\bar{H}_{3,k} = \mathbb{E}[H_{3,k} \mid H_{3,k-\kappa}^{k-1}]$ and where $\tilde{H}_{3,k}$ is a zero-mean, complex Gaussian random variable of variance $\epsilon_{3,\kappa}^2$ satisfying

$$\lim_{\kappa \rightarrow \infty} \epsilon_{3,\kappa}^2 = \epsilon_3^2 = \exp\left(\int_{-1/2}^{1/2} \log F_3'(\lambda) d\lambda\right). \quad (72)$$

We thus have

$$\begin{aligned} I(X_k, X_{r,k}; Y_k \mid H_{3,k-\kappa}^{k-1}) &= h(Y_k \mid H_{3,k-\kappa}^{k-1}) - h(Y_k \mid X_k, X_{r,k}, H_{3,k-\kappa}^{k-1}) \\ &\geq h(Y_k \mid H_{3,k-\kappa}^{k-1}, H_{3,k}, H_{2,k}, X_k, Z_k) - h(Y_k \mid X_k, X_{r,k}, \bar{H}_{3,k}) \\ &= h(H_{3,k} X_{r,k} \mid H_{3,k}) - h(\tilde{H}_{3,k} X_{r,k} + H_{2,k} X_k + Z_k \mid X_k, X_{r,k}) \end{aligned} \quad (73)$$

where the second step follows because conditioning does not increase entropy and because $\bar{H}_{3,k}$ is a function of $H_{3,k-\kappa}^{k-1}$; and the last step follows from the property of differential entropy under translation and because, conditioned on $H_{3,k}$, the random variable $H_{3,k} X_{r,k}$ is independent of $(H_{3,k-\kappa}^{k-1}, H_{2,k}, X_k, Z_k)$.

As above, we further lower-bound the RHS of (73) as

$$\begin{aligned} I(X_k, X_{r,k}; Y_k \mid H_{3,k-\kappa}^{k-1}) &\geq h(H_{3,k} X_{r,k} \mid H_{3,k}) - h(\tilde{H}_{3,k} X_{r,k} + H_{2,k} X_k + Z_k \mid X_k, X_{r,k}) \\ &= \mathbb{E}[\log |H_{3,k}|^2] + h(X_{r,k}) - \mathbb{E}[\log |X_{r,k}|^2] - h\left(\tilde{H}_{3,k} + H_{1,k} \frac{X_k}{X_{r,k}} + \frac{Z_k}{X_{r,k}} \mid X_k, X_{r,k}\right) \\ &= \log(\log \mathbf{A}_r^2 - \beta \log \mathbf{A}_r^2) - \gamma - 1 - \mathbb{E}\left[\log\left(\epsilon_{3,\kappa}^2 + \frac{|X_k|^2}{|X_{r,k}|^2} + \frac{\sigma^2}{|X_{r,k}|^2}\right)\right] \\ &\geq \log \log \mathbf{A}_r^2 + \log(1 - \beta) - \gamma - 1 - \log\left(\epsilon_{3,\kappa}^2 + \frac{\mathbf{A}^{2\alpha}}{\mathbf{A}_r^{2\beta}} + \frac{\sigma^2}{\mathbf{A}_r^{2\beta}}\right) \\ &= \log \log(\rho^2 \mathbf{A}^2) + \log(1 - \beta) - \gamma - 1 - \log\left(\epsilon_{3,\kappa}^2 + \rho^{-2\beta} \mathbf{A}^{-2(\beta-\alpha)} + \rho^{-2\beta} \frac{\sigma^2}{\mathbf{A}^{2\beta}}\right) \end{aligned} \quad (74)$$

where the second step follows from the behavior of differential entropy under scaling by a complex number; the third step follows by evaluating $\mathbb{E}[\log |H_{3,k}|^2] + h(X_{r,k}) - \mathbb{E}[\log |X_{r,k}|^2]$ and because, conditioned on $(X_k, X_{r,k})$, the random variable $\tilde{H}_{3,k} + H_{1,k} X_k / X_{r,k} + Z_k / X_{r,k}$ is complex Gaussian; the fourth step follows because, for our choice of input distribution (55) and (56), we have $|X_k|^2 \leq \mathbf{A}^{2\alpha}$ and $|X_{r,k}|^2 \geq \mathbf{A}_r^{2\beta}$ with probability one; and the last step follows from (8).

Combining (68)–(74), and noting that the RHS of (74) does not depend on k , we obtain

$$\begin{aligned} \frac{1}{n} (X_1^n, X_{r,1}^n; Y_1^n) &\geq \frac{n - \kappa}{n} \left[\log \log(\rho^2 \mathbf{A}^2) + \log(1 - \beta) - \gamma - 1 \right. \\ &\quad \left. - \log\left(\epsilon_{3,\kappa}^2 + \rho^{-2\beta} \mathbf{A}^{-2(\beta-\alpha)} + \rho^{-2\beta} \frac{\sigma^2}{\mathbf{A}^{2\beta}}\right) - \varepsilon_2(\text{SNR}, \kappa) \right] \end{aligned} \quad (75)$$

which tends to

$$\log \log(\rho^2 A^2) + \log(1 - \beta) - \gamma - 1 - \log\left(\epsilon_{3,\kappa}^2 + \rho^{-2\beta} A^{-2(\beta-\alpha)} + \rho^{-2\beta} \frac{\sigma^2}{A^{2\beta}}\right) - \varepsilon_2(\text{SNR}, \kappa)$$

as n tends to infinity. Using (70), and noting that for every $0 < \alpha < \beta < 1$

$$\begin{aligned} \lim_{\text{SNR} \rightarrow \infty} \log \log(\rho^2 A^2) - \log \log \text{SNR} &= 0 \\ \lim_{\text{SNR} \rightarrow \infty} \log\left(\epsilon_{3,\kappa}^2 + \rho^{-2\beta} A^{-2(\beta-\alpha)} + \rho^{-2\beta} \frac{\sigma^2}{A^{2\beta}}\right) &= \log \epsilon_{3,\kappa}^2 \end{aligned}$$

we obtain

$$\lim_{\text{SNR} \rightarrow \infty} \left\{ \lim_{n \rightarrow \infty} \frac{1}{n} (X_1^n, X_{r,1}^n; Y_1^n) - \log \log \text{SNR} \right\} \geq -1 - \gamma + \log \frac{1}{\epsilon_{3,\kappa}^2} + \log(1 - \beta) \quad (76)$$

which tends to

$$\lim_{\text{SNR} \rightarrow \infty} \left\{ \lim_{n \rightarrow \infty} \frac{1}{n} (X_1^n, X_{r,1}^n; Y_1^n) - \log \log \text{SNR} \right\} \geq -1 - \gamma + \log \frac{1}{\epsilon_3^2} + \log(1 - \beta) \quad (77)$$

as we let κ tend to infinity.

7.3 Maximizing Over α and β

It follows from (54), (67) and (77) that with a DF strategy, we can achieve the fading number

$$\chi \geq \min \left\{ -1 - \gamma + \log \frac{1}{\epsilon_1^2} + \log \alpha, -1 - \gamma + \log \frac{1}{\epsilon_3^2} + \log(1 - \beta) \right\} \quad (78)$$

for every $0 < \alpha < \beta < 1$. We prove Theorem 3 by maximizing over α and β . To this end, we first note that the optimal choice of α and β (in the sense that it achieves the supremum over $0 < \alpha < \beta < 1$) must satisfy $\alpha = \beta$, since otherwise we could either increase α or decrease β to obtain a higher achievable rate. We further note that the optimal $\alpha = \beta$ must satisfy

$$-1 - \gamma + \log \frac{1}{\epsilon_1^2} + \log \alpha = -1 - \gamma + \log \frac{1}{\epsilon_3^2} + \log(1 - \alpha). \quad (79)$$

Indeed, if we had

$$-1 - \gamma + \log \frac{1}{\epsilon_1^2} + \log \alpha < -1 - \gamma + \log \frac{1}{\epsilon_3^2} + \log(1 - \alpha) \quad (80)$$

then it would follow that

$$\min \left\{ -1 - \gamma + \log \frac{1}{\epsilon_1^2} + \log \alpha, -1 - \gamma + \log \frac{1}{\epsilon_3^2} + \log(1 - \alpha) \right\} = -1 - \gamma + \log \frac{1}{\epsilon_1^2} + \log \alpha \quad (81)$$

which could be increased by increasing α . In contrast, if we had

$$-1 - \gamma + \log \frac{1}{\epsilon_1^2} + \log \alpha > -1 - \gamma + \log \frac{1}{\epsilon_3^2} + \log(1 - \alpha) \quad (82)$$

then it would follow that

$$\min \left\{ -1 - \gamma + \log \frac{1}{\epsilon_1^2} + \log \alpha, -1 - \gamma + \log \frac{1}{\epsilon_3^2} + \log(1 - \alpha) \right\} = -1 - \gamma + \log \frac{1}{\epsilon_3^2} + \log(1 - \alpha) \quad (83)$$

which could be increased by decreasing α . Solving (79) yields

$$\alpha = \frac{\epsilon_1^2}{\epsilon_1^2 + \epsilon_3^2}. \quad (84)$$

Combining (84) with (78) proves Theorem 3.

8 Quantize Map and Forward

Recently, a strategy called *quantize-map-and-forward* was introduced by Avestimehr et al. [17]. They showed that the quantize-map-and-forward scheme achieves rates that are within a constant gap of the max-flow min-cut upper bound, where the gap depends on the number of relays but not on the channel parameters. For example, for the Gaussian relay channel with a single relay, and for the two-relay Gaussian diamond network, the gap is not more than one bit.

However, for the Gaussian relay channel with a single relay, rates that are within one bit of the max-flow min-cut upper bound can also be achieved by a decode-and-forward scheme [17, Th. 3.1]. We therefore believe that for the above fading relay channel, the quantize-map-and-forward scheme will give rise to communication rates that are comparable to the ones presented in Theorem 3. (For fading relay channels with more than one relay, the quantize-map-and-forward scheme may be superior to the decode-and-forward scheme.)

Indeed, if the link between the transmitter and the relay supports higher rates than the link between the relay and the receiver, then the decode-and-forward scheme achieves rates that are within one bit of the capacity (Corollary 4). If the link between the transmitter and the relay supports smaller rates than the link between the relay and the receiver, then the gap between the upper bound (16) and the lower bound (21) can be larger than one bit.

9 Conclusion

We studied the high-SNR asymptotic capacity of fading relay channels. We considered a noncoherent model where all terminals are aware of the statistics of the fading, but not of their realizations. We demonstrated that, if the link between the transmitter and the receiver supports higher communication rates than the link between the relay and the receiver, then at high SNR it is optimal to turn the relay off. We further demonstrated that if the link between the transmitter and the relay supports higher communication rates than the link between the relay and the receiver, then a decode-and-forward strategy achieves communication rates that are within one bit of the high-SNR capacity of the multiple-input single-output fading channel that results when the transmitter and the relay can cooperate.

A Proof of Proposition 5

To prove Proposition 5, we evaluate (54) for $\{X_k, k \in \mathbb{Z}\}$ and $\{X_{r,k}, k \in \mathbb{Z}\}$ being i.i.d., circularly-symmetric random variables, independent of each other and with

$$\log |X_k|^2 \sim \mathcal{U}([\alpha \log(\delta^2 A^2), \alpha \log A^2]), \quad k \in \mathbb{Z} \quad (85)$$

$$\log |X_{r,k}|^2 \sim \mathcal{U}([\log(\delta_r^2 A_r^2), \log A_r^2]), \quad k \in \mathbb{Z} \quad (86)$$

where $0 < \alpha, \delta, \delta_r < 1$. We shall first evaluate the first term on the RHS of (54), namely,

$$\lim_{n \rightarrow \infty} \frac{1}{n} I(X_1^n; Y_{r,1}^n | X_{r,1}^n) = \lim_{n \rightarrow \infty} \frac{1}{n} I(X_1^n; Y_{r,1}^n). \quad (87)$$

To lower-bound (87), we use [18, Prop. 4.1] with A^2 replaced by $A^{2\alpha}$, and with α replace by $\delta\alpha$. We thus obtain

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} I(X_1^n; Y_{r,1}^n) &\geq \log\left(\frac{\sigma^2}{\delta^\alpha A^{2\alpha}}\right) - \int_{-1/2}^{1/2} \log\left(F_1'(\lambda) + \frac{\sigma^2}{\delta^{2\alpha} A^{2\alpha}}\right) d\lambda \\ &\quad - \exp\left(\frac{\sigma^2 e}{\alpha \log(\frac{1}{\delta^2}) \delta^\alpha A^{2\alpha}}\right) \text{Ei}\left(-\frac{\sigma^2 e}{\alpha \log(\frac{1}{\delta^2}) \delta^\alpha A^{2\alpha}}\right) \\ &= \log\left(\frac{\sigma^{2(1-\alpha)}}{\delta^\alpha \text{SNR}^\alpha}\right) - \int_{-1/2}^{1/2} \log\left(F_1'(\lambda) + \frac{\sigma^{2(1-\alpha)}}{\delta^{2\alpha} \text{SNR}^\alpha}\right) d\lambda \\ &\quad - \exp\left(\frac{\sigma^{2(1-\alpha)} e}{\alpha \log(\frac{1}{\delta^2}) \delta^\alpha \text{SNR}^\alpha}\right) \text{Ei}\left(-\frac{\sigma^{2(1-\alpha)} e}{\alpha \log(\frac{1}{\delta^2}) \delta^\alpha \text{SNR}^\alpha}\right), \quad \text{SNR} > 0. \quad (88) \end{aligned}$$

We next lower-bound the second term on the RHS of (54), namely, $\lim_{n \rightarrow \infty} \frac{1}{n} I(X_1^n, X_{r,1}^n; Y_1^n)$. We use the chain rule for mutual information and a Cesàro-type theorem [12, Th. 4.2.3] to lower-bound

$$\begin{aligned}
& \lim_{n \rightarrow \infty} \frac{1}{n} I(X_1^n, X_{r,1}^n; Y_1^n) \\
&= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n I(X_k, X_{r,k}; Y_1^n \mid X_1^{k-1}, X_{r,1}^{k-1}) \\
&\geq \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n I(X_k, X_{r,k}; Y_1^k \mid X_1^{k-1}, X_{r,1}^{k-1}) \\
&\geq \lim_{k \rightarrow \infty} I(X_k, X_{r,k}; Y_1^k \mid X_1^{k-1}, X_{r,1}^{k-1}) \\
&\geq \lim_{k \rightarrow \infty} \inf I\left(X_k, X_{r,k}; Y_k, \left\{ \frac{Y_\ell}{X_{r,\ell}} \right\}_{\ell=1}^{k-1} \mid X_1^{k-1} = x_1^{k-1}, X_{r,1}^{k-1} = x_{r,1}^{k-1}\right) \tag{89}
\end{aligned}$$

where \lim denotes the *limit inferior*, and where the infimum in the last step is over all $(x_1^{k-1}, x_{r,1}^{k-1})$ satisfying

$$\delta^{2\alpha} \mathbf{A}^{2\alpha} \leq |x_\ell|^2 \leq \mathbf{A}^{2\alpha} \quad \text{and} \quad \delta_r^2 \mathbf{A}_r^2 \leq |x_{r,\ell}|^2 \leq \mathbf{A}_r^2, \quad \ell = 1, 2, \dots, k-1.$$

Here the second step follows because reducing observations does not increase mutual information.

We next show that the RHS of (89) is minimized when $|x_\ell|^2 = \mathbf{A}^{2\alpha}$ and $|x_{r,\ell}|^2 = \delta_r^2 \mathbf{A}_r^2$ for $\ell = 1, 2, \dots, k-1$. Indeed, conditioned on $(X_1^{k-1}, X_{r,1}^{k-1}) = (x_1^{k-1}, x_{r,1}^{k-1})$, the random variables $Y_\ell/x_{r,\ell}$ are given by

$$\frac{Y_\ell}{x_{r,\ell}} = H_{3,\ell} + H_{2,\ell} \frac{x_\ell}{x_{r,\ell}} + \frac{Z_\ell}{x_{r,\ell}} \stackrel{\mathcal{L}}{=} H_{3,\ell} + \left(\frac{x_\ell}{x_{r,\ell}} + \frac{\sigma^2}{x_{r,\ell}} \right) W_\ell, \quad \ell = 1, 2, \dots, k-1 \tag{90}$$

where $\{W_k, k \in \mathbb{Z}\}$ is a sequence of i.i.d., zero-mean, unit-variance, complex Gaussian random variables, and where $A \stackrel{\mathcal{L}}{=} B$ indicates that A and B have the same law. Here we use in the second step that $\{H_{2,k}, k \in \mathbb{Z}\}$ and $\{Z_k, k \in \mathbb{Z}\}$ are both sequences of i.i.d. Gaussian random variables. The second term on the RHS of (90) can be viewed as an additive-noise term. Thus, by choosing $|x_\ell|^2 = \mathbf{A}^{2\alpha}$ and $|x_{r,\ell}|^2 = \delta_r^2 \mathbf{A}_r^2$, the variance of the additive noise is maximized. By contradiction, it is easy to show that maximizing the variance of the additive noise minimizes the mutual information. Indeed, suppose that the noise that minimizes the mutual information is not the one with maximum variance. Then, we can add i.i.d. zero-mean Gaussian noise $\{U_k, k \in \mathbb{Z}\}$ to $Y_\ell/x_{r,\ell}$ such that $Y_\ell/x_{r,\ell} + U_\ell$ has the same distribution as $Y_\ell/x_{r,\ell}$ when $|x_\ell|^2 = \mathbf{A}^{2\alpha}$ and $|x_{r,\ell}|^2 = \delta_r^2 \mathbf{A}_r^2$. Since adding noise cannot increase the mutual information, it follows that the choice $|x_\ell|^2 = \mathbf{A}^{2\alpha}$ and $|x_{r,\ell}|^2 = \delta_r^2 \mathbf{A}_r^2$ minimizes the mutual information.

We thus obtain for the RHS of (89)

$$\begin{aligned}
& \lim_{k \rightarrow \infty} \inf I\left(X_k, X_{r,k}; Y_k, \left\{ \frac{Y_\ell}{X_{r,\ell}} \right\}_{\ell=1}^{k-1} \mid X_1^{k-1} = x_1^{k-1}, X_{r,1}^{k-1} = x_{r,1}^{k-1}\right) \\
&= \lim_{k \rightarrow \infty} I\left(X_k, X_{r,k}; Y_k, \left\{ H_{3,\ell} + \left(\frac{1}{\delta_r^2 \rho^2 \mathbf{A}^{2(1-\alpha)}} + \frac{\sigma^2}{\delta_r^2 \rho^2 \mathbf{A}^2} \right) W_\ell \right\}_{\ell=1}^{k-1}\right) \\
&= \lim_{k \rightarrow \infty} I\left(X_k, X_{r,k}; Y_k \mid \left\{ H_{3,\ell} + \left(\frac{1}{\delta_r^2 \rho^2 \mathbf{A}^{2(1-\alpha)}} + \frac{\sigma^2}{\delta_r^2 \rho^2 \mathbf{A}^2} \right) W_\ell \right\}_{\ell=1}^{k-1}\right) \tag{91}
\end{aligned}$$

where we have used that $\mathbf{A}_r = \rho \mathbf{A}$ (8). Here the first step follows because the joint law of

$$\left(X_k, X_{r,k}, Y_k, \left\{ H_{3,\ell} + \left(\frac{1}{\delta_r^2 \rho^2 \mathbf{A}^{2(1-\alpha)}} + \frac{\sigma^2}{\delta_r^2 \rho^2 \mathbf{A}^2} \right) W_\ell \right\}_{\ell=1}^{k-1} \right)$$

does not depend on $(x_1^{k-1}, x_{r,1}^{k-1})$; and the last step follows because the pair $(X_k, X_{r,k})$ is independent of $(\{H_{3,k}, k \in \mathbb{Z}\}, \{W_k, k \in \mathbb{Z}\})$.

We continue by expressing the mutual information as the difference of two differential entropies, i.e., we have

$$\begin{aligned} I\left(X_k, X_{r,k}; Y_k \middle| \left\{ H_{3,\ell} + \left(\frac{1}{\delta_r^2 \rho^2 \mathcal{A}^{2(1-\alpha)}} + \frac{\sigma^2}{\delta_r^2 \rho^2 \mathcal{A}^2} \right) W_\ell \right\}_{\ell=1}^{k-1} \right) \\ = h\left(Y_k \middle| \left\{ H_{3,\ell} + \left(\frac{1}{\delta_r^2 \rho^2 \mathcal{A}^{2(1-\alpha)}} + \frac{\sigma^2}{\delta_r^2 \rho^2 \mathcal{A}^2} \right) W_\ell \right\}_{\ell=1}^{k-1} \right) \\ - h\left(Y_k \middle| X_k, X_{r,k}, \left\{ H_{3,\ell} + \left(\frac{1}{\delta_r^2 \rho^2 \mathcal{A}^{2(1-\alpha)}} + \frac{\sigma^2}{\delta_r^2 \rho^2 \mathcal{A}^2} \right) W_\ell \right\}_{\ell=1}^{k-1} \right). \end{aligned} \quad (92)$$

For the second differential entropy, we have

$$\begin{aligned} h\left(Y_k \middle| X_k, X_{r,k}, \left\{ H_{3,\ell} + \left(\frac{1}{\delta_r^2 \rho^2 \mathcal{A}^{2(1-\alpha)}} + \frac{\sigma^2}{\delta_r^2 \rho^2 \mathcal{A}^2} \right) W_\ell \right\}_{\ell=1}^{k-1} \right) \\ = \mathbb{E}[\log |X_{r,k}|^2] \\ + h\left(H_{3,k} + H_{2,k} \frac{X_k}{X_{r,k}} + \frac{Z_k}{X_{r,k}} \middle| X_k, X_{r,k}, \left\{ H_{3,\ell} + \left(\frac{1}{\delta_r^2 \rho^2 \mathcal{A}^{2(1-\alpha)}} + \frac{\sigma^2}{\delta_r^2 \rho^2 \mathcal{A}^2} \right) W_\ell \right\}_{\ell=1}^{k-1} \right) \\ = \log(\delta_r \rho^2 \mathcal{A}^2) + \log \pi + 1 + \log(\epsilon_{3,k}^2(\xi) + \xi) \end{aligned} \quad (93)$$

where

$$\xi = \frac{1}{\delta_r^2 \rho^2 \mathcal{A}^{2(1-\alpha)}} + \frac{\sigma^2}{\delta_r^2 \rho^2 \mathcal{A}^2}$$

and where $\epsilon_{3,k}^2(\xi)$ denotes the mean-square error in predicting $H_{3,k}$ from $(H_{3,k-1} + \xi W_{k-1}), \dots, (H_{3,1} + \xi W_1)$. Note that [5, Sec. III]

$$\lim_{k \rightarrow \infty} \epsilon_{3,k}^2(\xi) = \exp\left(\int_{-1/2}^{1/2} \log(F'_3(\lambda) + \xi) d\lambda\right) - \xi. \quad (94)$$

The last step in (93) follows by evaluating $\mathbb{E}[\log |X_{r,k}|^2]$ and by noting that, conditioned on

$$\left(X_k, X_{r,k}, \left\{ H_{3,\ell} + \left(\frac{1}{\delta_r^2 \rho^2 \mathcal{A}^{2(1-\alpha)}} + \frac{\sigma^2}{\delta_r^2 \rho^2 \mathcal{A}^2} \right) W_\ell \right\}_{\ell=1}^{k-1} \right)$$

the random variable $H_{3,k} + H_{2,k} X_k / X_{r,k} + Z_k / X_{r,k}$ is complex Gaussian with variance $\epsilon_{3,k}^2(\xi) + \xi$.

For the first differential entropy on the RHS of (92), we have

$$\begin{aligned} h\left(Y_k \middle| \left\{ H_{3,\ell} + \left(\frac{1}{\delta_r^2 \rho^2 \mathcal{A}^{2(1-\alpha)}} + \frac{\sigma^2}{\delta_r^2 \rho^2 \mathcal{A}^2} \right) W_\ell \right\}_{\ell=1}^{k-1} \right) \\ \geq h(H_{3,k} X_{r,k} + H_{2,k} X_k + Z_k \mid H_{3,k}) \\ \geq \mathbb{E}\left[\log\left(e^{\log |H_{3,k}|^2 + h(X_{r,k})} + e^{h(H_{2,k} X_k)} + \pi e \sigma^2\right)\right] \end{aligned} \quad (95)$$

where the first step follows because conditioning does not increase entropy and because, conditional on $H_{3,k}$, the channel output Y_k is independent of

$$\left\{ H_{3,\ell} + \left(\frac{1}{\delta_r^2 \rho^2 \mathcal{A}^{2(1-\alpha)}} + \frac{\sigma^2}{\delta_r^2 \rho^2 \mathcal{A}^2} \right) W_\ell \right\}_{\ell=1}^{k-1};$$

and the second step follows from the Entropy Power Inequality [12, Th. 16.6.3] and from the property of differential entropy under scaling by a complex number. For the distribution (86) on $X_{r,k}$, we have

$$\begin{aligned} h(X_{r,k}) &= h(\log |X_{r,k}|^2) + \mathbb{E}[\log |X_{r,k}|^2] + \log \pi \\ &= \log\left(\log\left(\frac{1}{\delta_r^2}\right) \delta_r \mathcal{A}_r^2 \pi\right) \end{aligned} \quad (96)$$

where we use in the first step that $X_{r,k}$ is circularly-symmetric [3, Lemmas 6.15 & 6.16]. Similarly, the entropy $h(H_{2,k}X_k)$ is lower-bounded by

$$\begin{aligned} h(H_{2,k}X_k) &\geq h(H_{2,k}X_k \mid H_{2,k}) \\ &= \mathbb{E}[\log |H_{2,k}|^2] + h(X_k) \\ &= \log \left(e^{-\gamma} \alpha \log \left(\frac{1}{\delta^2} \right) \delta^\alpha A^{2\alpha} \pi \right) \end{aligned} \quad (97)$$

where we use that $\mathbb{E}[\log |H_{2,k}|^2] = -\gamma$ and $h(X_k) = h(\log |X_k|^2) + \mathbb{E}[\log |X_k|^2] + \log \pi$ [3, Lemmas 6.15 & 6.16]. Combining (96) and (97) with (95) yields

$$\begin{aligned} &h \left(Y_k \mid \left\{ H_{3,\ell} + \left(\frac{1}{\delta_r^2 \rho^2 A^{2(1-\alpha)}} + \frac{\sigma^2}{\delta_r^2 \rho^2 A^2} \right) W_\ell \right\}_{\ell=1}^{k-1} \right) \\ &\geq \mathbb{E} \left[\log \left(|H_{3,k}|^2 \log \left(\frac{1}{\delta_r^2} \right) \delta_r \rho^2 A^2 \pi + e^{-\gamma} \alpha \log \left(\frac{1}{\delta^2} \right) \delta^\alpha A^{2\alpha} \pi + \pi e \sigma^2 \right) \right] \\ &= \log \pi + \log \log \frac{1}{\delta_r^2} + \log(\delta_r \rho^2 A^2) + \mathbb{E} \left[\log \left(|H_{3,k}|^2 + \frac{e^{-\gamma} \alpha \log \left(\frac{1}{\delta^2} \right) \delta^\alpha A^{2\alpha} + e \sigma^2}{\log \left(\frac{1}{\delta_r^2} \right) \delta_r \rho^2 A^2} \right) \right] \\ &= \log \pi + \log \log \frac{1}{\delta_r^2} + \log(\delta_r \rho^2 A^2) + \log \left(\frac{e^{-\gamma} \alpha \log \left(\frac{1}{\delta^2} \right) \delta^\alpha A^{2\alpha} + e \sigma^2}{\log \left(\frac{1}{\delta_r^2} \right) \delta_r \rho^2 A^2} \right) \\ &\quad - \exp \left(\frac{e^{-\gamma} \alpha \log \left(\frac{1}{\delta^2} \right) \delta^\alpha A^{2\alpha} + e \sigma^2}{\log \left(\frac{1}{\delta_r^2} \right) \delta_r \rho^2 A^2} \right) \text{Ei} \left(- \frac{e^{-\gamma} \alpha \log \left(\frac{1}{\delta^2} \right) \delta^\alpha A^{2\alpha} + e \sigma^2}{\log \left(\frac{1}{\delta_r^2} \right) \delta_r \rho^2 A^2} \right) \end{aligned} \quad (98)$$

where the last step follows by noting that $|H_{3,k}|^2$ has an exponential distribution for which the expectation can be computed using [22, Sec. 4.337]. Combining (98) and (93) yields

$$\begin{aligned} &I \left(X_k, X_{r,k}; Y_k \mid \left\{ H_{3,\ell} + \left(\frac{1}{\delta_r^2 \rho^2 A^{2(1-\alpha)}} + \frac{\sigma^2}{\delta_r^2 \rho^2 A^2} \right) W_\ell \right\}_{\ell=1}^{k-1} \right) \\ &\geq \log \log \frac{1}{\delta_r^2} - 1 + \log \left(\frac{e^{-\gamma} \alpha \log \left(\frac{1}{\delta^2} \right) \delta^\alpha A^{2\alpha} + e \sigma^2}{\log \left(\frac{1}{\delta_r^2} \right) \delta_r \rho^2 A^2} \right) - \log(\epsilon_{r,k}^2(\xi) + \xi) \\ &\quad - \exp \left(\frac{e^{-\gamma} \alpha \log \left(\frac{1}{\delta^2} \right) \delta^\alpha A^{2\alpha} + e \sigma^2}{\log \left(\frac{1}{\delta_r^2} \right) \delta_r \rho^2 A^2} \right) \text{Ei} \left(- \frac{e^{-\gamma} \alpha \log \left(\frac{1}{\delta^2} \right) \delta^\alpha A^{2\alpha} + e \sigma^2}{\log \left(\frac{1}{\delta_r^2} \right) \delta_r \rho^2 A^2} \right) \end{aligned} \quad (99)$$

where ξ is given in (93). By (94), the RHS of (99) tends to

$$\begin{aligned} &\log \log \frac{1}{\delta_r^2} - 1 + \log \left(\frac{e^{-\gamma} \alpha \log \left(\frac{1}{\delta^2} \right) \delta^\alpha A^{2\alpha} + e \sigma^2}{\log \left(\frac{1}{\delta_r^2} \right) \delta_r \rho^2 A^2} \right) \\ &\quad - \int_{-1/2}^{1/2} \log \left(F'_3(\lambda) + \frac{1}{\delta_r^2 \rho^2 A^{2(1-\alpha)}} + \frac{\sigma^2}{\delta_r^2 \rho^2 A^2} \right) d\lambda \\ &\quad - \exp \left(\frac{e^{-\gamma} \alpha \log \left(\frac{1}{\delta^2} \right) \delta^\alpha A^{2\alpha} + e \sigma^2}{\log \left(\frac{1}{\delta_r^2} \right) \delta_r \rho^2 A^2} \right) \text{Ei} \left(- \frac{e^{-\gamma} \alpha \log \left(\frac{1}{\delta^2} \right) \delta^\alpha A^{2\alpha} + e \sigma^2}{\log \left(\frac{1}{\delta_r^2} \right) \delta_r \rho^2 A^2} \right) \end{aligned} \quad (100)$$

as k tends to infinity. It thus follows from (89)–(100) that

$$\begin{aligned} &\lim_{n \rightarrow \infty} \frac{1}{n} I(X_1^n, X_{r,1}^n; Y_1^n) \\ &\geq \log \log \frac{1}{\delta_r^2} - 1 + \log \left(\frac{e^{-\gamma} \alpha \log \left(\frac{1}{\delta^2} \right) \delta^\alpha A^{2\alpha} + e \sigma^2}{\log \left(\frac{1}{\delta_r^2} \right) \delta_r \rho^2 A^2} \right) \\ &\quad - \int_{-1/2}^{1/2} \log \left(F'_3(\lambda) + \frac{1}{\delta_r^2 \rho^2 A^{2(1-\alpha)}} + \frac{\sigma^2}{\delta_r^2 \rho^2 A^2} \right) d\lambda \end{aligned}$$

$$\begin{aligned}
& - \exp\left(\frac{e^{-\gamma}\alpha \log\left(\frac{1}{\delta^2}\right)\delta^\alpha A^{2\alpha} + e\sigma^2}{\log\left(\frac{1}{\delta_r^2}\right)\delta_r \rho^2 A^2}\right) \text{Ei}\left(-\frac{e^{-\gamma}\alpha \log\left(\frac{1}{\delta^2}\right)\delta^\alpha A^{2\alpha} + e\sigma^2}{\log\left(\frac{1}{\delta_r^2}\right)\delta_r \rho^2 A^2}\right) \\
& = \log \log \frac{1}{\delta_r^2} - 1 + \log\left(\frac{e^{-\gamma}\alpha \log\left(\frac{1}{\delta^2}\right)\delta^\alpha \text{SNR}^\alpha \sigma^{2(\alpha-1)} + e}{\log\left(\frac{1}{\delta_r^2}\right)\delta_r \rho^2 \text{SNR}}\right) \\
& - \int_{-1/2}^{1/2} \log\left(F'_3(\lambda) + \frac{\sigma^{2(\alpha-1)}}{\delta_r^2 \rho^2 \text{SNR}^{1-\alpha}} + \frac{1}{\delta_r^2 \rho^2 \text{SNR}}\right) d\lambda \\
& - \exp\left(\frac{e^{-\gamma}\alpha \log\left(\frac{1}{\delta^2}\right)\delta^\alpha \text{SNR}^\alpha \sigma^{2(\alpha-1)} + e}{\log\left(\frac{1}{\delta_r^2}\right)\delta_r \rho^2 \text{SNR}}\right) \text{Ei}\left(-\frac{e^{-\gamma}\alpha \log\left(\frac{1}{\delta^2}\right)\delta^\alpha \text{SNR}^\alpha \sigma^{2(\alpha-1)} + e}{\log\left(\frac{1}{\delta_r^2}\right)\delta_r \rho^2 \text{SNR}}\right) \quad (101)
\end{aligned}$$

for $\text{SNR} > 0$. Combining (88), (101), and (54), and maximizing over $0 < \delta, \alpha, \delta_r < 1$, proves Proposition 5.

B Proof of Proposition 6

The proof of Proposition 6 is an extension of the DF strategy analyzed in [15] to channels with memory. As in the memoryless case, it uses a technique called *block-Markov superposition encoding*. Most steps of the proof for memoryless channels [15] can be easily extended to channels with memory by defining the set of typical sequences via entropy rates rather than via entropies, cf. (103). The main difference is that for memoryless channels, the events (114) and (115) are independent of each other, whereas for channels with memory, these events are dependent. Consequently, we obtain a third term on the RHS of (116), for which we need to show that its exponent equals to zero, cf. (119). Below we give a detailed proof (cf. [23, Ch. 9]).

Codebook construction: Encoding is performed in $B + 1$ blocks of n symbols. For each block, we generate a separate codebook. That is, we fix some distribution $P_{X, X_r}(\cdot)$ and some rate \tilde{R} . Then, for every block b , $b = 1, \dots, B + 1$ the codebook of the relay is constructed by drawing $e^{n\tilde{R}}$ codewords $x_{r,1}^n(v; b)$, $v = 1, \dots, e^{n\tilde{R}}$ i.i.d. according to the distribution $P_{X_r}(\cdot)$. As for the codebook of the transmitter, for every $v = 1, \dots, e^{n\tilde{R}}$ we generate $e^{n\tilde{R}}$ codewords $x_1^n(w, v; b)$, $w = 1, \dots, e^{n\tilde{R}}$ independently according to the conditional distribution $P_{X|X_r}(\cdot)$, i.e., we draw each symbol $x_k(w, v; b)$ according to $P_{X|X_r}(\cdot | x_{r,k}(v; b))$.

In the proof, we assume that $P_{X, X_r}(\cdot)$ is such that the random variables $(X_1^n, X_{r,1}^n)$ have a probability density function, which implies that also $(X_1^n, X_{r,1}^n, Y_{r,1}^n, Y_1^n)$ have a probability density function. (We shall denote the probability density function of the random variable A by $f_A(\cdot)$.) The case where $P_{X, X_r}(\cdot)$ does not allow for a probability density function $f_{X_1^n, X_{r,1}^n, Y_{r,1}^n, Y_1^n}(\cdot)$ can be treated by partitioning the sample spaces of the channel inputs and outputs into a finite collection of mutually exclusive events, and by studying the resulting discrete problem following the steps below. (To this end, we need to replace the *differential* entropy rates in the definition of jointly typical sequences (103) with entropy rates.) The result follows then by taking the supremum over all partitions, cf. [24, Sec. 2.5].

Transmitter: The message m to be transmitted is divided into B equally-sized blocks m_1, \dots, m_B of $e^{n\tilde{R}}$ nats each. In block b , $b = 1, \dots, B + 1$ the transmitter sends out the code-word $x_1^n(m_b, m_{b-1}; b)$, where we define $m_0 = m_{B+1} = 1$.

Relay: After the transmission of block b is completed, the relay has observed the sequence of outputs $y_{r,1}^n(b)$. The relay tries to find an $m_{r,b}$ such that

$$(x_1^n(m_{r,b}, \hat{m}_{r,b-1}; b), x_{r,1}^n(\hat{m}_{r,b-1}; b), y_{r,1}^n(b)) \in \mathcal{A}_\epsilon(X_1^n, X_{r,1}^n, Y_{r,1}^n) \quad (102)$$

where $\hat{m}_{r,b-1}$ denotes the relay's estimate of the message for block $b-1$, and where $\mathcal{A}_\epsilon(X_1^n, X_{r,1}^n, Y_{r,1}^n)$ denotes the *set of jointly typical sequences* with respect to $P_{X_1^n, X_{r,1}^n, Y_{r,1}^n}(\cdot)$. That is

$$\begin{aligned}
\mathcal{A}_\epsilon(A_{1,1}^n, \dots, A_{\tau,1}^n) \triangleq & \left\{ a_{\mathcal{I},1}^n \in \mathbb{C}^{n|\mathcal{I}|}, \forall \mathcal{I} \subseteq \{1, \dots, \tau\} : \right. \\
& \left. \left| -\frac{1}{n} \log f_{A_{\mathcal{I},1}^n}(a_{\mathcal{I},1}^n) - h(\{A_{\mathcal{I},k}\}) \right| < \epsilon \right\} \quad (103)
\end{aligned}$$

where $a_{\mathcal{I},1}^n$ denotes the set of sequences $a_{t,1}^n$ with $t \in \mathcal{I}$; $|\mathcal{I}|$ denotes the cardinality of the set \mathcal{I} ; and $h(\{A_{\mathcal{I},k}\})$ denotes the entropy rate of the random processes $\{A_{t,k}, k \in \mathbb{Z}\}$, $t \in \mathcal{I}$, i.e.,

$$h(\{A_{\mathcal{I},k}\}) \triangleq \lim_{n \rightarrow \infty} \frac{h(A_{\mathcal{I},1}^n)}{n}.$$

If one or more $m_{r,b}$ can be found satisfying (102), then the relay chooses one of them, calls this choice $\hat{m}_{r,b}$, and transmits $x_{r,1}^n(\hat{m}_{r,b}; b+1)$ in the subsequent block. If no such $\hat{m}_{r,b}$ is found, then the relay sets $\hat{m}_{r,b} = 1$ and transmits $x_{r,1}^n(1; b+1)$ in the subsequent block.

Receiver: After block b , $b = 2, \dots, B+1$ the receiver has observed the outputs $y_1^n(b-1)$ and $y_1^n(b)$. It tries to find an m_{b-1} such that

$$(x_1^n(m_{b-1}, \hat{m}_{b-2}; b-1), x_{r,1}^n(\hat{m}_{b-2}; b-1), y_1^n(b-1)) \in \mathcal{A}_\epsilon(X_1^n, X_{r,1}^n, Y_1^n) \quad (104)$$

$$(x_{r,1}^n(m_{b-1}; b), y_1^n(b)) \in \mathcal{A}_\epsilon(X_{r,1}^n, Y_1^n) \quad (105)$$

where \hat{m}_{b-2} is the receiver's estimate of m_{b-2} . If one or more such m_{b-1} are found, then the receiver chooses one of them and calls this choice \hat{m}_{b-1} . If no such m_{b-1} is found, then the receiver sets $\hat{m}_{b-1} = 1$.

Analysis: For each block b , $b = 1, \dots, B+1$, let $\mathcal{E}_{r,b}^{(0)}$ denote the event that the relay cannot find an $m_{r,b}$ that satisfies (102), and let $\mathcal{E}_{r,b}^{(+)}$ denote the event that the relay chooses an $\hat{m}_{r,b} \neq m_b$ satisfying (102). Similarly, let $\mathcal{E}_b^{(0)}$ denote the event that the receiver cannot find an m_{b-1} that satisfies (104) and (105), and let $\mathcal{E}_b^{(+)}$ denote the event that the receiver chooses an $\hat{m}_{b-1} \neq m_{b-1}$ satisfying (104) and (105). Finally, let \mathcal{F}_{b-1} be the event that no errors have been made up to block b . The probability of error is upper-bounded by

$$\begin{aligned} \Pr(\text{error}) &\leq \Pr\left(\bigcup_{b=1}^B (\mathcal{E}_{r,b}^{(0)} \cup \mathcal{E}_{r,b}^{(+)}) \cup \bigcup_{b=2}^{B+1} (\mathcal{E}_b^{(0)} \cup \mathcal{E}_b^{(+)})\right) \\ &= \Pr(\mathcal{E}_{r,1}^{(0)} \cup \mathcal{E}_{r,1}^{(+)}) + \sum_{b=2}^B \Pr\left((\mathcal{E}_{r,b}^{(0)} \cup \mathcal{E}_{r,b}^{(+)}) \cup (\mathcal{E}_b^{(0)} \cup \mathcal{E}_b^{(+)}) \cap \mathcal{F}_{b-1}\right) \\ &\quad + \Pr(\mathcal{E}_{B+1}^{(0)} \cup \mathcal{E}_{B+1}^{(+)} \cap \mathcal{F}_B). \end{aligned} \quad (106)$$

It follows that we can analyze every block separately by assuming that no errors were made in the previous blocks. The overall probability of error is then upper-bounded by $(B+1)$ times the maximum error probability of each block.

Suppose that no errors occurred up to block b . Then, it follows from the union bound that the error probability in block b is upper-bounded by

$$\begin{aligned} &\Pr\left((\mathcal{E}_{r,b}^{(0)} \cup \mathcal{E}_{r,b}^{(+)}) \cup (\mathcal{E}_b^{(0)} \cup \mathcal{E}_b^{(+)}) \cap \mathcal{F}_{b-1}\right) \\ &\leq \Pr(\mathcal{E}_{r,b}^{(0)} \cap \mathcal{F}_{b-1}) + \Pr(\mathcal{E}_{r,b}^{(+)} \cap \mathcal{F}_{b-1}) + \Pr(\mathcal{E}_b^{(0)} \cap \mathcal{F}_{b-1}) + \Pr(\mathcal{E}_b^{(+)} \cap \mathcal{F}_{b-1}). \end{aligned} \quad (107)$$

In order to upper-bound (107), we first note that, for a given (m_b, m_{b-1}) , the process $\{(X_k, X_{r,k}), k \in \mathbb{Z}\}$ is i.i.d. and jointly independent of the stationary and ergodic, complex Gaussian fading processes $\{H_{\ell,k}, k \in \mathbb{Z}\}$, $\ell = 1, 2, 3$ and of the i.i.d. Gaussian noise processes $\{Z_{r,k}, k \in \mathbb{Z}\}$ and $\{Z_k, k \in \mathbb{Z}\}$, which implies that the process $\{(X_k, X_{r,k}, Y_{r,k}, Y_k), k \in \mathbb{Z}\}$ is jointly stationary and ergodic. It thus follows from the Shannon-McMillan-Breiman theorem [25, Th. 2] that

$$\lim_{n \rightarrow \infty} \Pr\left((X_1^n(m_b, m_{b-1}; b), X_{r,1}^n(m_{b-1}; b), Y_{r,1}^n(b)) \in \mathcal{A}_\epsilon(X_1^n, X_{r,1}^n, Y_{r,1}^n)\right) = 1 \quad (108)$$

$$\lim_{n \rightarrow \infty} \Pr\left((X_1^n(m_{b-1}, m_{b-2}; b-1), X_{r,1}^n(m_{b-2}; b-1), Y_1^n(b-1)) \in \mathcal{A}_\epsilon(X_1^n, X_{r,1}^n, Y_1^n)\right) = 1 \quad (109)$$

$$\lim_{n \rightarrow \infty} \Pr\left((X_{r,1}^n(m_{b-1}; b), Y_1^n(b)) \in \mathcal{A}_\epsilon(X_{r,1}^n, Y_1^n)\right) = 1. \quad (110)$$

This implies that

$$\Pr(\mathcal{E}_{r,b}^{(0)} \cap \mathcal{F}_{b-1}) \quad \text{and} \quad \Pr(\mathcal{E}_b^{(0)} \cap \mathcal{F}_{b-1})$$

both tend to zero as n tends to infinity.

We continue with the error event $(\mathcal{E}_{r,b}^{(+)} \cap \mathcal{F}_{b-1})$. This event occurs if the relay finds an $\hat{m}_{r,b} \neq m_b$ such that

$$(x_1^n(\hat{m}_{r,b}, m_{b-1}; b), x_{r,1}^n(m_{b-1}; b), y_{r,1}^n(b)) \in \mathcal{A}_\epsilon(X_1^n, X_{r,1}^n, Y_{r,1}^n).$$

Since we have $\hat{m}_b \neq m_b$, it follows that $(X_1^n(\hat{m}_{r,b}, m_{r,b-1}; b), X_{r,1}^n(m_{b-1}; b), Y_{r,1}^n(b))$ is distributed according to $P_{X_{r,1}^n}(\cdot)P_{X_1^n|X_{r,1}^n}(\cdot)P_{Y_{r,1}^n|X_{r,1}^n}(\cdot)$. Extending [12, Th. 14.2.3] to channels with memory³ yields that, for every $\hat{m}_{r,b} \neq m_b$, we have

$$\begin{aligned} \Pr\left((X_1^n(\hat{m}_{r,b}, m_{b-1}; b), X_{r,1}^n(\hat{m}_{r,b}; b), Y_{r,1}^n(b)) \in \mathcal{A}_\epsilon(X_1^n, X_{r,1}^n, Y_{r,1}^n)\right) \\ \leq \exp\left(-n\left(\lim_{n \rightarrow \infty} \frac{1}{n} I(X_1^n; Y_{r,1}^n \mid X_{r,1}^n) - 6\epsilon\right)\right). \end{aligned} \quad (111)$$

It thus follows from the union bound that

$$\begin{aligned} \Pr(\mathcal{E}_{r,b}^{(+)} \cap \mathcal{F}_{b-1}) &\leq \sum_{\hat{m}_{r,b} \neq m_b} \exp\left(-n\left(\lim_{n \rightarrow \infty} \frac{1}{n} I(X_1^n; Y_{r,1}^n \mid X_{r,1}^n) - 6\epsilon\right)\right) \\ &\leq \exp\left(n\left(\tilde{R} - \lim_{n \rightarrow \infty} \frac{1}{n} I(X_1^n; Y_{r,1}^n \mid X_{r,1}^n) + 6\epsilon\right)\right). \end{aligned} \quad (112)$$

This implies that, if

$$\tilde{R} < \lim_{n \rightarrow \infty} \frac{1}{n} I(X_1^n; Y_{r,1}^n \mid X_{r,1}^n) - 6\epsilon \quad (113)$$

then the probability $\Pr(\mathcal{E}_{r,b}^{(+)} \cap \mathcal{F}_{b-1})$ vanishes as n tends to infinity.

We finally consider the error event $(\mathcal{E}_b^{(+)} \cap \mathcal{F}_{b-1})$. This event occurs if the receiver finds an $\hat{m}_{b-1} \neq m_{b-1}$ that satisfies

$$(x_1^n(\hat{m}_{b-1}, m_{b-2}; b-1), x_{r,1}^n(m_{b-2}; b-1), y_1^n(b-1)) \in \mathcal{A}_\epsilon(X_1^n, X_{r,1}^n, Y_1^n) \quad (114)$$

$$(x_{r,1}^n(\hat{m}_{b-1}; b), y_1^n(b)) \in \mathcal{A}_\epsilon(X_{r,1}^n, Y_1^n). \quad (115)$$

Since we have $\hat{m}_{b-1} \neq m_{b-1}$, it follows that

$$(X_1^n(\hat{m}_{b-1}, m_{b-2}; b-1), X_{r,1}^n(m_{b-2}; b-1), Y_1^n(b-1), X_{r,1}^n(\hat{m}_{b-1}; b), Y_1^n(b))$$

is distributed according to

$$P_{X_{r,1}^n(b-1)}(\cdot)P_{X_1^n(b-1)|X_{r,1}^n(b-1)}(\cdot)P_{Y_1^n(b-1)|X_{r,1}^n(b-1)}(\cdot)P_{X_{r,1}^n(b)}(\cdot)P_{Y_1^n(b)|X_{r,1}^n(b-1), Y_1^n(b-1)}(\cdot)$$

where the argument after the random vector indicates whether the vector belongs to block b or $(b-1)$. Extending [12, Ths. 14.2.1 & 14.2.3] to the above channel, we obtain for every $\hat{m}_{b-1} \neq m_{b-1}$

$$\begin{aligned} \Pr((114) \text{ \& } (115) \text{ are satisfied}) \\ \leq \exp\left(-n\left(\lim_{n \rightarrow \infty} \frac{1}{n} I(X_1^n(b-1); Y_1^n(b-1) \mid X_{r,1}^n(b-1)) + \lim_{n \rightarrow \infty} \frac{1}{n} I(X_{r,1}^n(b); Y_1^n(b)) - 10\epsilon\right)\right) \\ \times \exp\left(n\left(\lim_{n \rightarrow \infty} \frac{1}{n} I(Y_1^n(b); X_{r,1}^n(b-1), Y_1^n(b-1))\right)\right). \end{aligned} \quad (116)$$

Since the codebook construction does not depend on the block b , and since the channel is stationary, it follows that

$$\lim_{n \rightarrow \infty} \frac{1}{n} I(X_1^n(b-1); Y_1^n(b-1) \mid X_{r,1}^n(b-1)) = \lim_{n \rightarrow \infty} \frac{1}{n} I(X_1^n; Y_1^n \mid X_{r,1}^n) \quad (117)$$

³To this end, we need to replace the entropies in the proof of [12, Th. 14.2.3] by the corresponding differential entropy rates.

and

$$\lim_{n \rightarrow \infty} \frac{1}{n} I(X_{r,1}^n(b); Y_1^n(b)) = \lim_{n \rightarrow \infty} \frac{1}{n} I(X_{r,1}^n; Y_1^n) \quad (118)$$

do not depend on b . We next show that

$$\lim_{n \rightarrow \infty} \frac{1}{n} I(Y_1^n(b); X_{r,1}^n(b-1), Y_1^n(b-1)) = 0. \quad (119)$$

To this end, we first note that, by the stationarity of $(Y_1^n(b), X_{r,1}^n(b-1), Y_1^n(b-1))$, we have

$$I(Y_1^n(b); X_{r,1}^n(b-1), Y_1^n(b-1)) = I(Y_1^n; X_{r,-n+1}^0, Y_{-n+1}^0). \quad (120)$$

This can be upper-bounded by

$$\begin{aligned} I(Y_1^n; X_{r,-n+1}^0, Y_{-n+1}^0) &\leq I(\{H_{2,\ell} X_\ell + H_{3,\ell} X_{r,\ell}\}_{\ell=1}^n; X_{r,-n+1}^0, \{H_{2,\ell} X_\ell + H_{3,\ell} X_{r,\ell}\}_{\ell=-n+1}^0) \\ &\leq I(H_{2,1}^n, H_{3,1}^n, X_1^n, X_{r,1}^n; H_{2,-n+1}^0, H_{3,-n+1}^0, X_{-n+1}^0, X_{r,-n+1}^0) \\ &= I(H_{2,1}^n, H_{3,1}^n; H_{2,-n+1}^0, H_{3,-n+1}^0) \\ &= I(H_{2,1}^n; H_{2,-n+1}^0) + I(H_{3,1}^n; H_{3,-n+1}^0) \\ &\leq I(H_{2,1}^n; H_{2,-\infty}^0) + I(H_{3,1}^n; H_{3,-\infty}^0) \end{aligned} \quad (121)$$

where the first two steps follow from the Data Processing Inequality [12, Th. 2.8.1]; the third step follows because the processes $\{X_k, k \in \mathbb{Z}\}$ and $\{X_{r,k}, k \in \mathbb{Z}\}$ are i.i.d. and independent of $\{H_{2,k}, k \in \mathbb{Z}\}$ and $\{H_{3,k}, k \in \mathbb{Z}\}$; the fourth step follows because the processes $\{H_{2,k}, k \in \mathbb{Z}\}$ and $\{H_{3,k}, k \in \mathbb{Z}\}$ are independent; and the last step follows because adding observations does not decrease mutual information. It follows by the chain rule for mutual information and by Cesàro's mean [12, Th. 4.2.3] that

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} I(H_{2,1}^n; H_{2,-\infty}^0) &= \lim_{k \rightarrow \infty} I(H_{2,k}; H_{2,-\infty}^0 \mid H_{2,1}^{k-1}) \\ &= \lim_{k \rightarrow \infty} h(H_{2,k} \mid H_{2,1}^{k-1}) - \lim_{k \rightarrow \infty} h(H_{2,k} \mid H_{2,-\infty}^{k-1}) \\ &= h(\{H_{2,k}\}) - h(\{H_{2,k}\}) \\ &= 0 \end{aligned} \quad (122)$$

where the third step follows from the stationarity of $\{H_{2,k}, k \in \mathbb{Z}\}$ [12, Th. 4.2.1]. In the same way, it can be shown that

$$\lim_{n \rightarrow \infty} \frac{1}{n} I(H_{3,1}^n; H_{3,-\infty}^0) = 0. \quad (123)$$

Combining (120)–(123) proves (119). We thus obtain from (116)–(119) that

$$\begin{aligned} &\Pr((114) \text{ \& } (115) \text{ are satisfied}) \\ &\leq \exp\left(-n\left(\lim_{n \rightarrow \infty} \frac{1}{n} I(X_1^n; Y_1^n \mid X_{r,1}^n) + \lim_{n \rightarrow \infty} \frac{1}{n} I(X_{r,1}^n; Y_1^n) - 10\epsilon\right)\right). \end{aligned} \quad (124)$$

By the union bound, it follows that

$$\begin{aligned} &\Pr(\mathcal{E}_b^{(+)} \cap \mathcal{F}_{b-1}) \\ &\leq \sum_{\hat{m}_{b-1} \neq m_{b-1}} \exp\left(-n\left(\lim_{n \rightarrow \infty} \frac{1}{n} I(X_1^n; Y_1^n \mid X_{r,1}^n) + \lim_{n \rightarrow \infty} \frac{1}{n} I(X_{r,1}^n; Y_1^n) - 10\epsilon\right)\right) \\ &\leq \exp\left(n\left(\tilde{R} - \lim_{n \rightarrow \infty} \frac{1}{n} I(X_1^n; Y_1^n \mid X_{r,1}^n) + \lim_{n \rightarrow \infty} \frac{1}{n} I(X_{r,1}^n; Y_1^n) + 10\epsilon\right)\right) \\ &= \exp\left(n\left(\tilde{R} - \lim_{n \rightarrow \infty} \frac{1}{n} I(X_1^n, X_{r,1}^n; Y_1^n) + 10\epsilon\right)\right) \end{aligned} \quad (125)$$

which implies that $\Pr(\mathcal{E}_b^{(+)} \cap \mathcal{F}_{b-1})$ vanishes as n tends to infinity, provided that

$$\tilde{R} < \lim_{n \rightarrow \infty} \frac{1}{n} I(X_1^n, X_{r,1}^n; Y_1^n) - 10\epsilon. \quad (126)$$

It thus follows from (126) and (113) that for every block b the rate

$$\tilde{R} = \min \left\{ \lim_{n \rightarrow \infty} \frac{1}{n} I(X_1^n, Y_{r,1}^n \mid X_{r,1}^n), \lim_{n \rightarrow \infty} \frac{1}{n} I(X_1^n, X_{r,1}^n; Y_1^n) \right\} \quad (127)$$

is achievable. Consequently, the complete message m can be transmitted over $B+1$ blocks with an overall rate

$$R = \frac{n\tilde{R}B}{n(B+1)} = \frac{B\tilde{R}}{B+1}. \quad (128)$$

By letting B tend to infinity, it follows that for every product distribution on $(X_1^n, X_{r,1}^n)$, i.e.,

$$P_{X_1^n, X_{r,1}^n}(\cdot) = \prod_{k=1}^n P_{X, X_r}(\cdot)$$

we can achieve the rate

$$R = \min \left\{ \lim_{n \rightarrow \infty} \frac{1}{n} I(X_1^n, Y_{r,1}^n \mid X_{r,1}^n), \lim_{n \rightarrow \infty} \frac{1}{n} I(X_1^n, X_{r,1}^n; Y_1^n) \right\}.$$

This proves Proposition 6.

C The limit of $\varepsilon_2(\text{SNR}, \kappa)$

In the following we show that $\varepsilon_2(\text{SNR}, \kappa)$ tends to zero as SNR tends to infinity. The proof follows along the same lines as the proof in [3, Appendix IX].

We first note that $\varepsilon_2(\text{SNR}, \kappa) \geq 0$. Thus, it suffices to show that $\varepsilon_2(\text{SNR}, \kappa) \leq o(1)$. We have

$$\begin{aligned} \varepsilon_2(\text{SNR}, \kappa) &= I(X_k, X_{r,k}; H_{3,k}^{k-1} \mid Y_{k-\kappa}^k, X_{k-\kappa}^{k-1}, X_{r,k-\kappa}^{k-1}) \\ &= h(H_{3,k-\kappa}^{k-1} \mid Y_{k-\kappa}^k, X_{k-\kappa}^{k-1}, X_{r,k-\kappa}^{k-1}) - h(H_{3,k-\kappa}^{k-1} \mid Y_{k-\kappa}^k, X_{k-\kappa}^k, X_{r,k-\kappa}^k) \\ &\leq h(H_{3,k-\kappa}^{k-1} \mid Y_{k-\kappa}^{k-1}, X_{k-\kappa}^{k-1}, X_{r,k-\kappa}^{k-1}) - h(H_{3,k-\kappa}^{k-1} \mid Y_{k-\kappa}^k, X_{k-\kappa}^k, X_{r,k-\kappa}^k, H_{3,k}) \\ &= h(H_{3,k-\kappa}^{k-1} \mid Y_{k-\kappa}^{k-1}, X_{k-\kappa}^{k-1}, X_{r,k-\kappa}^{k-1}) - h(H_{3,k-\kappa}^{k-1} \mid Y_{k-\kappa}^{k-1}, X_{k-\kappa}^{k-1}, X_{r,k-\kappa}^{k-1}, H_{3,k}) \\ &= I(H_{3,k-\kappa}^{k-1}; H_{3,k} \mid Y_{k-\kappa}^{k-1}, X_{k-\kappa}^{k-1}, X_{r,k-\kappa}^{k-1}) \end{aligned} \quad (129)$$

where the third step follows because conditioning cannot increase entropy; and the fourth step follows because, conditioned on $(Y_{k-\kappa}^{k-1}, X_{k-\kappa}^{k-1}, X_{r,k-\kappa}^{k-1}, H_{3,k})$, the fading coefficients $H_{3,k-\kappa}^{k-1}$ are independent of $(Y_k, X_k, X_{r,k})$.

Expressing the mutual information $I(H_{3,k-\kappa}^{k-1}; H_{3,k} \mid Y_{k-\kappa}^{k-1}, X_{k-\kappa}^{k-1}, X_{r,k-\kappa}^{k-1})$ as the difference of two conditional differential entropies of $H_{3,k}$ yields

$$\begin{aligned} \varepsilon_2(\text{SNR}, \kappa) &\leq h(H_{3,k} \mid Y_{k-\kappa}^{k-1}, X_{k-\kappa}^{k-1}, X_{r,k-\kappa}^{k-1}) - h(H_{3,k} \mid Y_{k-\kappa}^{k-1}, H_{3,k-\kappa}^{k-1}, X_{k-\kappa}^{k-1}, X_{r,k-\kappa}^{k-1}) \\ &= h(H_{3,k} \mid Y_{k-\kappa}^{k-1}, X_{k-\kappa}^{k-1}, X_{r,k-\kappa}^{k-1}) - h(H_{3,k} \mid H_{3,k-\kappa}^{k-1}) \\ &= h\left(H_{3,k} \mid \left\{ H_{2,\ell} \frac{X_\ell}{X_{r,\ell}} + H_{3,\ell} + \frac{Z_\ell}{X_{r,\ell}} \right\}_{\ell=k-\kappa}^{k-1}, X_{k-\kappa}^{k-1}, X_{r,k-\kappa}^{k-1}\right) - h(H_{3,k} \mid H_{3,k-\kappa}^{k-1}) \\ &\leq h\left(H_{3,k} \mid \{H_{3,\ell} + \zeta W_\ell\}_{\ell=k-\kappa}^{k-1}\right) - h(H_{3,k} \mid H_{3,k-\kappa}^{k-1}) \end{aligned} \quad (130)$$

where $\{W_k, k \in \mathbb{Z}\}$ is a sequence of i.i.d., zero-mean, unit-variance, circularly-symmetric, complex Gaussian random variables, and where

$$\zeta \triangleq \frac{A^{2\alpha}}{A_r^{2\beta}} + \frac{\sigma^2}{A_r^{2\beta}} = \rho^{-2\beta} A^{-2(\beta-\alpha)} + \rho^{-2\beta} \frac{\sigma^2}{A^{2\beta}}.$$

The second step in (130) follows because, conditioned on $H_{3,k-\kappa}^{k-1}$, the present fading $H_{3,k}$ is independent of $(X_{k-\kappa}^{k-1}, X_{r,k-\kappa}^{k-1}, Y_{k-\kappa}^{k-1})$; and the last step in (130) follows because the first differential entropy is maximized for $|X_\ell|^2 = A^{2\alpha}$ and $|X_{r,\ell}|^2 = A_r^{2\beta}$, in which case

$$\left\{ H_{2,\ell} \frac{X_\ell}{X_{r,\ell}} + H_{3,\ell} + \frac{Z_\ell}{X_{r,\ell}} \right\}_{\ell=k-\kappa}^{k-1}$$

has the same law as $\{H_{3,\ell} + \zeta W_\ell\}_{\ell=k-\kappa}^{k-1}$. Noting that

$$\begin{aligned} & h(H_{3,k} \mid \{H_{3,\ell} + \zeta W_\ell\}_{\ell=k-\kappa}^{k-1}) - h(H_{3,k} \mid H_{3,k-\kappa}^{k-1}) \\ &= h(H_{3,k}, \{H_{3,\ell} + \zeta W_\ell\}_{\ell=k-\kappa}^{k-1}) - h(H_{3,k}, H_{3,k-\kappa}^{k-1}) - I(\{H_{3,\ell} + \zeta W_\ell\}_{\ell=k-\kappa}^{k-1}; W_{k-\kappa}^{k-1}) \\ &\leq h(H_{3,k}, \{H_{3,\ell} + \zeta W_\ell\}_{\ell=k-\kappa}^{k-1}) - h(H_{3,k}, H_{3,k-\kappa}^{k-1}) \end{aligned} \quad (131)$$

we obtain

$$\varepsilon_2(\text{SNR}, \kappa) \leq h(H_{3,k}, \{H_{3,\ell} + \zeta W_\ell\}_{\ell=k-\kappa}^{k-1}) - h(H_{3,k}, H_{3,k-\kappa}^{k-1}). \quad (132)$$

The claim follows now by [3, Lemma 6.11], which states that if $\mathbf{H} \in \mathbb{C}^\nu$ is a random vector of finite Frobenius norm and finite differential entropy, and if $\mathbf{W} \in \mathbb{C}^\nu$ is a Gaussian random vector that is independent of \mathbf{H} , then

$$\lim_{\sigma^2 \rightarrow 0} \{h(\mathbf{H} + \sigma^2 \mathbf{W}) - h(\mathbf{H})\} = 0.$$

Acknowledgment

Fruitful discussions with Amos Lapidoth and Alfonso Martinez are gratefully acknowledged.

References

- [1] G. Kramer, M. Gastpar, and P. Gupta, "Cooperative strategies and capacity theorems for relay networks," *IEEE Transactions on Information Theory*, vol. 51, no. 9, pp. 3037–3063, September 2005.
- [2] B. Wang, J. Zhang, and A. Høst-Madsen, "On the capacity of MIMO relay channels," *IEEE Transactions on Information Theory*, vol. 51, no. 1, pp. 29–43, January 2005.
- [3] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat fading channels," *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2426–2467, October 2003.
- [4] T. H. E. Ericson, "A Gaussian channel with slow fading," *IEEE Transactions on Information Theory*, vol. 16, no. 3, pp. 353–355, May 1970.
- [5] A. Lapidoth, "On the asymptotic capacity of stationary Gaussian fading channels," *IEEE Transactions on Information Theory*, vol. 51, no. 2, pp. 437–446, February 2005.
- [6] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Transactions on Information Theory*, vol. 45, no. 1, pp. 139–157, January 1999.
- [7] L. Zheng and D. N. C. Tse, "Communicating on the Grassmann manifold: a geometric approach to the non coherent multiple-antenna channel," *IEEE Transactions on Information Theory*, vol. 48, no. 2, pp. 359–383, February 2002.
- [8] J. Doob, *Stochastic Processes*. John Wiley & Sons, 1990.
- [9] A. Lapidoth and S. M. Moser, "The fading number of single-input multiple-output fading channels with memory," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 437–453, February 2006.

- [10] A. Lapidoth, "On the high SNR capacity of stationary Gaussian fading channels," in *Proceedings Forty-First Allerton Conference on Communication, Control and Computing*, Allerton House, Monticello, Illinois, October 1–3, 2003.
- [11] T. Koch and A. Lapidoth, "Degrees of freedom in non-coherent stationary MIMO fading channels," in *Proceedings Winter School on Coding and Information Theory*, Bratislava, Slovakia, February 20–25, 2005.
- [12] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 1st ed. John Wiley & Sons, 1991.
- [13] T. Koch and A. Lapidoth, "The fading number and degrees of freedom in non-coherent MIMO fading channels: a peace pipe," in *Proceedings IEEE International Symposium on Information Theory (ISIT)*, Adelaide, Australia, September 4–9, 2005.
- [14] T. Koch, "On the asymptotic capacity of multiple-input single-output fading channels with memory," Master's thesis, Signal and Information Processing Laboratory, ETH Zurich, Switzerland, April 2004, supervised by Prof. Dr. Amos Lapidoth.
- [15] T. M. Cover and A. A. El Gamal, "Capacity theorems for the relay channel," *IEEE Transactions on Information Theory*, vol. 25, no. 5, pp. 572–584, September 1979.
- [16] R. Etkin and D. Tse, "Degrees of freedom in some underspread MIMO fading channels," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1576–1608, April 2006.
- [17] A. S. Avestimehr, S. N. Diggavi, and D. N. C. Tse, "Wireless network information flow: A deterministic approach," *IEEE Transactions on Information Theory*, vol. 57, no. 4, pp. 1872–1905, April 2011.
- [18] T. Koch, "On heating up and fading in communication channels," Ph.D. dissertation, Swiss Federal Institute of Technology, Zurich, May 2009, Diss. ETH No. 18339.
- [19] V. Sethuraman, B. Hajek, and K. Narayanan, "Capacity bounds for noncoherent fading channels with a peak constraint," in *Proceedings IEEE International Symposium on Information Theory (ISIT)*, Adelaide, Australia, September 4–9, 2005.
- [20] V. Sethuraman, L. Wang, B. Hajek, and A. Lapidoth, "Low-SNR capacity of noncoherent fading channels," *IEEE Transactions on Information Theory*, vol. 55, no. 4, pp. 1555–1574, April 2009.
- [21] N. Wiener and P. Masani, "The prediction theory of multivariate stochastic processes I," *Acta Mathematica*, vol. 98, pp. 111–150, November 1957.
- [22] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 6th ed., A. Jeffrey, Ed. Academic Press, San Diego, 2000.
- [23] G. Kramer, *Topics in Multi-User Information Theory*. Foundations and Trends in Communications and Information Theory, 2007, vol. 4, no. 4–5, pp. 265–444.
- [24] R. G. Gallager, *Information Theory and Reliable Communication*. John Wiley & Sons, 1968.
- [25] P. H. Algoet and T. M. Cover, "A sandwich proof of the Shannon-McMillan-Breiman theorem," *The Annals of Probability*, vol. 16, no. 2, pp. 899–909, 1988.